„Zoophilologica. Polish Journal of Animal Studies"
Nr 1 (15)/2025, pp. 1–43
ISSN 2451-3849
DOI: https://doi.org/10.31261/ZOOPHILOLOGICA.2025.15.15

VARIA

Zbigniew Słuszkiewicz
https://orcid.org/0000-0001-7734-8484
Uniwersytet im. KEN w Krakowie

Of Rats and Men III
A Pragmatist Reconstruction of Advanced Agency *via* the Active Inference

О крысах и людях III
Прагматистская реконструкция развитой агентности посредством активного вывода

Абстракт

Заключительная статья триптиха *О крысах и людях* подводит итоги всех его частей, цель которых: (1) поставить под сомнение основные положения традиционного философского дискурса о разуме человека и разуме живых существ, не принадлежащих к человеческому роду; (2) продемонстрировать, как инструменты «прагматистского поворота» могут углубить понимание агентности животных; и (3) эмпирически обосновать градуалистские представления о моральной агентности, такие как концепция «морального агента» Марка Роулендса. Эти цели достигаются путем изучения характеристик развитой агентности. Продолжая тему, представленную во второй части триптиха *О крысах и людях*, автор ссылается на теорию активного вывода (англ. *Active Inference*, AIN), чтобы выявить, каким образом процессы контрфактического вывода и самоочевидности способствуют ее формированию. В этом контексте теоретический анализ Якоба Хохви

Of Rats and Men III
A Pragmatist Reconstruction of Advanced Agency *via* the Active Inference

Abstract

The final part of the triptych (*Of Rats and Men*) synthesises the entire series, which aims to (1) challenge selected foundational assumptions of traditional philosophical discourse on human and nonhuman minds, (2) demonstrate how the tools of "the pragmatic turn" can deepen our understanding of animal agency, and (3) provide *a posteriori* support for gradualist categories of moral agency, such as Mark Rowlands's concept of "the moral subject." We pursue these goals by exploring the characteristic features of an advanced form of agency. Building on the themes introduced in *Of Rats and Men II*, we employ the theory of active inference (AIN) to determine how the processes of counterfactual inferencing and self-evidencing foster its emergence. In this context, Jakob Hohwy's theoretical analysis is juxtaposed with experimental results on rodents, suggesting that the capacity for experiencing self-evidencing at a psychophysical level falls within the scope of the sense of agency (SoA) and may extend beyond our species.

соотносится с результатами экспериментов на грызунах, указывающими на то, что способность к самоочевидности на психофизиологическом уровне становится компонентом чувства агентности (англ. *Sense of Agency*, SoA) и не является исключительно свойством человека. Для изучения этого явления предлагается интерпретация «временного связывания» (разработанного Антонеллой Трамасере и Колином Алленом) с использованием теории активного вывода. С целью продемонстрировать онтологическую целостность концепции крыс как существ, способных достичь уровня когнитивно развитой агентности, автор соотносится ее с концепцией «когнитивных световых конусов» Майкла Левина.

Ключевые слова: активный вывод, самоочевидность, чувство агентности, моральный агент, крысы

To investigate this phenomenon, we propose an AIN-based interpretation of Antonella Tramacere and Colin Allen's "temporal binding" experiment. Finally, we position the concept of rats as beings capable of achieving a cognitively sophisticated level of agency within Michael Levin's "cognitive light cones" framework to illustrate its ontological coherence.

Keywords: active inference, self-evidencing, sense of agency, moral subject, rats

I think therefore I am if I am what I think.[1]
I am, therefore, I think.[2]
Karl Friston

The primary aims of the *Of Rats and Men* triptych, of which this article constitutes the final part, are threefold: (1) to highlight questionable assumptions embedded within the fundamentals of philosophical discourse concerning non-human minds; (2) to demonstrate how the conceptual resources of the "pragmatic turn" can contribute to a better understanding of agency in animals; and (3) to strengthen – in a manner compatible with theoretical and empirical advancements – gradualist conceptions of moral subjectivity, such as the "moral subject" category proposed by Mark Rowlands.

In the first part (*Of Rats and Men I*),[3] we critically examined the concept of free will (FW) through the lens of classical and contemporary pragmatism, questioning

[1] Karl J. Friston, "Embodied Inference: Or 'I Think Therefore I Am, If I Am What I Think,'" in *The Implications of Embodiment: Cognition and Communication*, eds. Wolfgang Tschacher and Claudia Bergomi (Charlottesville: Imprint Academic, 2011), 892011.

[2] Karl J. Friston, "I Am Therefore I Think," in *The Unconscious: A Bridge Between Psychoanalysis and Cognitive Neuroscience*, eds. Marianne Leuzinger-Bohleber, Simon Arnold, and Mark Solms (London: Routledge, 2017), 113.

[3] Zbigniew Słuszkiewicz, "Of Rats and Men I: A Pragmatist Take on the Concept of Free Will as a Challenge to the Human-Animal Dichotomy," *Zoophilologica. Polish Journal of Animal Studies* vol. 2, no. 14 (2024): 1–43, accessed December 12, 2024, https://doi.org/10.31261/ZOOPHILOLOGICA.2024.14.12.

both its philosophical assumptions and empirical viability. We underscored how this concept perpetuates the human-animal dichotomy and how it impedes gradualist models of moral agency. Concurrently, we assessed the adequacy of arguments related to FW – such as the "space of reasons," "the ability to do otherwise," and "second-order desires – that are traditionally invoked in philosophical discussions on the qualitative differences between humans and other animals. We concluded by advocating the abandonment of the habitual recourse to FW in favour of a more inclusive approach to lived experience, namely, agency – a concept that modern philosophical (e.g., rationalist, analytical) and scientific (e.g., behavioural, associationist and notably biological) traditions have heretofore withheld from animals.

In the second part (*Of Rats and Men II*), we set in motion a reconstruction of the concept of agency by adopting a "bottom-up" approach grounded in first principles. We harnessed the conceptual tools of unorthodox cognitive theories, infused with pragmatist insights – the so-called "pragmatic turn." Our interpretation of basal biological agency has been shaped by integrating resources from the embodied cognition paradigm (chiefly enactivism), the free-energy principle (FEP), and the predictive processing theory (PP). This trajectory was further illuminated by the processual perspective of the metaphysics of biology and delineated by the first four (out of eight) naturalistic criteria for agency (NA) posited by Henry Potter and Kevin Mitchell. Ultimately, we argued that such a framework facilitates a grasp of intentional concepts – such as beliefs, desires, and reason-goal directedness – that can be successfully applied to the explanation of animal behaviour, serving as tools for effectively executing the homeodynamic imperative.

Building on these foundations, in the third part, we continue the project by exploring the properties necessary for the emergence of advanced forms of agency. We contend that our approach enables the identification of features of agency characteristics in cognitively advanced organisms, such as mammals (in our examples – rodents). We begin by applying the remaining four non-reductionist criteria for recognising agency proposed by Potter and Mitchell to brained animals, reinforcing these criteria with insight from processual metaphysics. We then introduce the active inference (AIN) theory, focusing on those aspects that emphasise the role of counterfactual inference and the psychophysical manifestations of self-evidencing in the phenomenon under question. We juxtapose Jakob Hohwy's theoretical analysis of these properties with empirical findings from experiments with rodents, which, in accordance with his criteria, plausibly reveal the presence of cognitive mechanisms that enhance agentive functions in these animals.

Subsequently, we briefly discuss research on the sense of agency (SoA) before introducing an experimental project authored by Antonella Tramacere and Colin Allen. The project is designed to investigate "temporal binding" in animals – a phenomenon discovered within the SoA research program. We propose interpreting it through AIN conceptual lenses to verify the occurrence of self-evidencing at the psychophysical level in organisms endowed with cerebral cortex. Lastly, we position our concept of agency, exemplified by rats, within the broader theoretical framework of Michael Levin's "cognitive light cones" to accentuate its ontological coherence. The article concludes with a brief discussion of our approach's philosophical consequences.

Final Steps of the Stroll from Agency Causation to Advanced Agency[4]

In the previous paper, we built upon our depiction of the first four criteria for "ontologically primitive causal power" – basal agency – demonstrating their compatibility with the free-energy principle (FEP) and the processual metaphysics of biology.[5] These are thermodynamic autonomy, persistence, endogenous activity, and holistic integration (see *Of Rats and Men II*). When an organism meets them, one can justifiably claim that it exhibits a basic form of causal agency – that is, it is capable of goal-directed actions based on reasons. At this point, we adduce the subsequent criteria and support them with the AIN tools, the workings of which become particularly evident at the level of neuronal architecture.[6]

The fifth criterion in Potter and Mitchell's dimensional list of standards for naturalised agency is low-level indeterminacy. It posits that a certain degree of unpredictability should exist even at the lowest level of system operations, thereby fostering the emergence of new, unforeseen capacities. Potter and Mitchell invoke here the idea of degrees to which the operation of higher structures in the system's hierarchy cannot be fully accounted for, solely by reference to microstructural

[4] With this title, we continue following Jakob von Uexküll's path.

[5] See Henry D. Potter and Kevin J. Mitchell, "Naturalising Agent Causation," *Entropy (Basel, Switzerland)*, vol. 24, no. 4 (2022): 1–8, accessed November 12, 2023, https://doi.org/10.3390/e24040472.

[6] These features already exist at "lower" organisational levels, and we must emphasise here that we are far from a neurocentric fetishisation of the brain. As Levin and Rouleau convincingly argue, no fundamental functional difference exists between neurons and other body cells. Neurons also do not possess any exclusive abilities or properties. Cognitive processes attributed to neuronal networks have their precursors outside of them, including problem-solving, navigation, communication, cooperation, learning, and memory. The fundamental claim of the approach they (and we) defend is the idea of continuity or homologies across different scales, where organisational principles precede substrates, see Nicolas Rouleau and Michael Levin, "Brains and Where Else? Mapping Theories of Consciousness to Unconventional Embodiments," *OSF Preprints*, 2025, 1–32, accessed January 16, 2025, https://doi.org/10.31234/osf.io/va5mk.

levels.[7] Indeterminacy is similarly expressed at the brain level. The system statistically extracts conspicuous features from the background (sampled noisy data) through its sensory surfaces. Signals arriving from the lower levels of the processing hierarchy may also gain perceived primary (felt) meaning if they deviate from the system's expectations embedded in prediction patterns.[8] Animal goal-oriented movement is steered (albeit not forced) by the activation of higher-order neural coalitions – shaped via probabilistic inference (interpretation, meaning-making).[9] At the brain–body–world level, these processes enable an animal to establish meaningful relations with her surroundings.[10] Inference thus influences actions through normative, but not determinative, calibration of those predicted features of the outside world that are either beneficial or detrimental to organismal homeodynamics. Orientation towards information-seeking (foraging or epistemic actions) provides an animal with a means of valuation. Through iterated experiences, she discerns and represents those aspects of the world that are salient to her – they acquire fixed meanings.[11]

[7] Unpredictability already exists at the quantum level, yet it is not necessarily an "objective indeterminacy" but rather probabilistic unpredictability from the perspective of a (bounded) observer.

[8] See, Mark Solms, *The Hidden Spring: A Journey to the Source of Consciousness* (New York: W. W. Norton & Company, 2021).

[9] The FEP framework abandons the logocentric approach to meaning, viewing it instead as a key ability of living systems to orient themselves in their environment, e.g., by using frames of reference. See Chris Fields and Michael Levin, "How Do Living Systems Create Meaning?," *Philosophies*, vol. 5, no. 4 (2020): 1–24, accessed January 20, 2025, https://doi.org/10.3390/philosophies5040036.

[10] We decided to refer to an animal as a feminine pronoun to emphasise that a living being is not a "thing." Within the FEP framework, a living being is a cognising system composed of subsets (defined in FEP as multiple, codependent, hierarchically-nested Markov blankets) tracking each other. Markov blanket separates inside from outside on many scales – from neurons to multicellular agents. A single blanket is a kind of veil (a boundary in statistical sense) encompassing sensory, internal and active states. A single cell consisting of at least one Markov blanket possesses sparse connections with the world through the sensory surface, enabling it to track external states and thus adjust to them or act (e.g. approach, avoid, i.e., *act for a reason*). From this perspective, the living being is a dynamical system – an active agent – her Markov-blanketed subsystems are metaphorically referred to as "a thing," see Maxwell Ramstead, "The Free Energy Principle – a Precis," *Dialectical Systems* (2023), accessed November 12, 2023, https://doi.org/10.5281/zenodo.10014870; Michael Kirchhoff et al., "The Markov Blankets of Life: Autonomy, Active Inference and the Free Energy Principle," *Journal of The Royal Society Interface*, vol. 15, no. 138 (2018), 20170792, accessed November 12, 2023, https://doi.org/10.1098/rsif.2017.0792.

[11] In the pragmatist depiction, "the meaning of a thing" is what an object or event (something that stands out) affords to an agent through experience. In James Gibson's ecological notion of affordance, there is an animal with a certain kind of body and brain coupled with the environment, which possesses certain physical structures and energy patterns. That environment will afford ways of interacting. In other words, "the meaning of something" is about how engagement through action impacts the experience and what it enacts or brings forth for the experience given a situation. "Meaningful information" is what features of the surroundings have been selected by an organism to act upon for a concrete purpose in virtue of its usefulness for successful persistence, see Philip Ball, *How Life*

As experience accumulates, some of these – such as structural and action-oriented representations or affordances (i.e., opportunities for action) – become integrated into the content of her internal states.[12] At the implementation level, internal content manifests through the activation of complex neuronal assemblies that expedite categorisation. Ultimately, by imbuing categories with subjective meaning, these processes facilitate the formation of situated natural concepts (Bayesian beliefs or simulators) that enable her to get a normative grip (in the Heideggerian sense) on certain aspects of the learned structure of the external world.[13] In this way, the fixed meanings of perceived situational contexts become predictable reasons for an animal's actions. The contentfulness of these phenomena follows from the fact that in the form of distinctive patterns of neuronally encoded probability distributions – formed *via* synaptic activity, neuronal connectivity, and structure entailed by the system's dynamics – fixed meanings are about something, whereas "aboutness" is traditionally considered the defining feature of intentionality.[14]

Works: A User's Guide to the New Biology (London: Picador, 2024), 349–351. For Deweyan's take on "meaning," see Mark L. Johnson and Jay Schulkin, *Mind in Nature: John Dewey, Cognitive Science, and a Naturalistic Philosophy for Living* (Cambridge, Massachusetts: The MIT Press, 2023), 50–53. We agree with such characterisation except for the Gibsonian rejection of inference. Sensorimotor coupling can also be successfully captured with this term.Within the AIN framework affordances are *a priori* high probability beliefs about hidden states, see Adam Linson et al., "The Active Inference Approach to Ecological Perception: General Information Dynamics for Natural and Artificial Embodied Cognition," *Frontiers in Robotics and AI*, vol. 5 (2018): 1–22, accessed January 17, 2025, https://www.frontiersin.org/articles/10.3389/frobt.2018.00021.

[12] Clearly, we are not adopting a notion of representation in the classical sense as a thorough and complete account of what is being represented – would it be a mental image, a three-dimensional model, a structure of predicates, or a configuration of amodal symbols "stored" in long-term memory. Instead, we are referring to representation in the dynamic and grounded sense as a consequence of the activation of neural networks and circuits. In this context, x represents y when x conveys information about y that enables goal-directed interaction with y by utilising that information. See Lawrence W. Barsalou, "Can Cognition Be Reduced to Action?," in *The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science*, eds. Andreas K. Engel, Karl J. Friston, and Danica Kragic (Cambridge, Massachusetts: The MIT Press, 2016), 81–96.

[13] See Lawrence W. Barsalou, "Continuity of the Conceptual System across Species," *Trends in Cognitive Sciences*, vol. 9, no. 7 (2005): 309–311, accessed September 20, 2023, https://doi.org/10.1016/j.tics.2005.05.003; Lawrence W. Barsalou, "Simulation, Situated Conceptualization, and Prediction," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1521 (2009): 1281–1289, accessed September 22, 2023, https://doi.org/10.1098/rstb.2008.0319. It is worth mentioning here that from the FEP perspective, concepts are not viewed as static representations but as dynamic capabilities of an agent to make specific conceptual distinctions using knowledge structurally encoded in her physical structure. As such, concepts are grounded and inherently normative, as they can be erroneously employed, see Wing Yi So, Karl J. Friston, and Victorita Neacsu, "The Inherent Normativity of Concepts," *Minds and Machines*, vol. 34, no. 4 (2024): 1–21, accessed January 10, 2025, https://doi.org/10.1007/s11023-024-09697-7.

[14] See Franz Brentano, *Psychologia z empirycznego punktu widzenia*, trans. Włodzimierz Galewicz (Warszawa: PWN, 1874/2006). See also Hans van Hateren, "Constructing a Naturalistic Theory

The following three criteria are pretty straightforward and do not require extensive elaboration. Multiple realisability, the sixth criterion, refers to the system's ability to achieve the same macroscopic outcomes through diverse microstructures. In other words, the relational variety among the system's components should enable the execution of similar goals in different ways, thereby enhancing the animal's flexibility (adaptability).[15] In the context of neural architecture, this principle again highlights the necessity for online functional interpretation of data arriving from lower levels of processing – a task executed by higher-level structures.

The seventh criterion, historicity, addresses the importance of considering not only "how" a given process proceeds (its mechanism) but also "why" it proceeds in one way rather than another. Evolutionary and ontogenetic histories enable us to determine why certain states of affairs are considered "good" and others "bad" from the perspective of a given organism. By positing historicity, researchers emphasise the crucial role of individual experience (e.g., learning, development) and evolutionary history (e.g., natural selection, epigenetics) in shaping an agent's current and future operating capabilities. Lastly, the eighth criterion – agent-level normativity – indicates the creature's need to act in accordance with internal goals, which are shaped by partly inherited and partly experientially formed meanings as the basis for reasons for action. At the highest levels of organisation, within organism-environment interplay, categories of intentionality proper can jump into play because complex actions (or policies) rest on complex reasons and distant goals. These operate on

of Intentionality," *Philosophia*, vol. 49 (2021), accessed September 23, 2023, https://doi.org/10.1007/s11406-020-00255-w. FEP provides a formal language for describing the "aboutness" of internal representations and, thus, a tool for a naturalistic understanding of intentionality as the ability of biological systems to interpret their own sensory streams as ontologies (representational structures). Within organism-centred functionalism, the external geometry of motion in the internal phase space of phenotypic states (not just neuronal states) separated from the world by a Markov blanket can be interpreted as a statistical parametrisation of probability distributions for external states. This dual geometry of information includes anticipatory hypotheses about the (internal and external) hidden causes of sensory states (these have properties of representational content). This means that all information (beliefs) about the system's environment is "compressed" in its internal states. Changes in internal states describe the distance between beliefs corresponding to updating the system's probabilistic inferences. In this way, anticipatory internal processes become the carrier of semantic content – the form of "knowledge" of the system about its environment, encoded in its structure and internal dynamics, see, Maxwell J. D. Ramstead, Karl J. Friston, and Inês Hipólito, "Is the Free-Energy Principle a Formal Theory of Semantics? From Variational Density Dynamics to Neural and Phenotypic Representations," *Entropy*, vol. 22, no. 8 (2020): 1–29, accessed March 15, 2025, https://doi.org/10.3390/e22080889.

[15] This feature is related on every level of organisation to William James' definition of intelligence, which has been picked up by empirical researchers embracing the AIN framework. James defined intelligence as the ability to reach the same goal in multiple ways; see Patrick McMillen and Michael Levin, "Collective Intelligence: A Unifying Concept for Integrating Biology Across Scales and Substrates," *Communications Biology*, vol. 7, no. 1 (2024): 5, accessed August 1, 2024, https://doi.org/10.1038/s42003-024-06037-4.

beliefs, desires, and intentions, which could be mere functional properties for neurons or neural networks, whereas "for an agent as a locus of meaning," they become cognitive skills.[16] Thereupon, the propensity of an agent to operate coherently – under internally established norms or goals – should enable the assessment of her actions as purposeful.[17] This naturalised account of how, which, and why biotic organisms might be regarded as acting upon their reasons whilst remaining consistent with modern biology, recognises nervous systems as control systems (or "cognitive glue") of diverse levels of causal power,[18] operating within organisms comprehended holistically in concordance with the embodied approach. It is vital to note that, under this view, it is not parts of the system that cause change; rather, only the system immersed in its niche, conceived as a whole, can be considered an acting agent.

Advanced Cognitive Agents – the Comparative Cognitive Science View

Recapitulating the above: as the meanings of stimuli became separated from the immediacy of re-action through multistage processing and feedback-loops within and between brain networks, the capacity to combine and deploy concepts (i.e., internal representations) to act emerges.[19] From this point on, distinctive patterns of neural activation (more broadly, sets of phenotypical states that gravitate towards particular "attractor states") not only entail pragmatic consequences manifested in the brain–body's actions. They also attain semantic content.[20] In other words, the activation of a particular neuronal population (assembly of neurons) structurally represents something in the world that is affectively meaningful for

[16] Potter and Mitchell, "Naturalising Agent Causation," 13.

[17] Potter and Mitchell, "Naturalising Agent Causation," 9–14. Notwithstanding our sympathy towards Potter and Mitchell's standards for the naturalised agency, we challenge the views of the latter scholar on FW. Mitchell defends FW by invoking the notion of "the space of reasons" and suggests the validity of the higher-order thought theory (HOT); see Kevin J. Mitchell, *Free Agents: How Evolution Gave Us Free Will*, first edition (Princeton Oxford: Princeton University Press, 2023), 276–282. This was recognised as fallacy by Rowlands, who termed it *the miracle-of-the-meta*; see Mark Rowlands, *Can Animals Be Moral?* (Oxford University Press, 2012), 169–190. We have argued against the classical notion "of the space of reasons" in *Of Rats and Men I*.

[18] See Michael Levin, "Bioelectric Networks: The Cognitive Glue Enabling Evolutionary Scaling from Physiology to Mind," *Animal Cognition* (2023), accessed September 10, 2023, https://doi.org/10.1007/s10071-023-01780-3.

[19] Potter and Mitchell, "Naturalising Agent Causation," 10.

[20] See footnotes 8, and 12.

the agent in whom it is activated.[21] Animals equipped with such neural organisation become true cognisers. As a result, such creatures gain the means to construct maps of the surroundings (initially physical, spatiotemporal, and later social) and to track changes emerging in their recurrently fluctuating environments.[22] The further development of the increasingly complex (mammalian) neocortex, primarily *via* the multiplication of existing microstructures called cortical columns,[23] along with the integration and consolidation of data from many modalities, enhances their internal models with knowledge about objects in the world and the causal relations between them, about properties, categories, associations, and regularity patterns.[24]

Due to the extensive robustness of control structures, some species gradually expand behavioural plasticity, coordination abilities, and the capacity to optimise adaptability (learning) not only on the basis of evolutionary processes but primarily through the accumulation of experiences gathered throughout their lives. Mammals and birds are those forms of life that have developed independently – through evolutionary convergence – particularly rich behavioural repertoire, to the extent that their actions exhibit second-order openness.[25] These organisms are endowed

[21] See, Potter and Mitchell, "Naturalising Agent Causation," 19.

[22] One neural implementation level, the basis for many of these capabilities can be found in cognitive maps (predominantly in the hippocampus and entorhinal cortex) predicted by Edward Tolman, see Jeff Hawkins, *A Thousand Brains: A New Theory of Intelligence* (New York: Hachette Book Group, 2021). Cognitive maps enable agents to perform inferential processes in spatiotemporal domain but also social and abstract spaces (see Jeffery A. Dusek and Howard Eichenbaum, "The Hippocampus and Memory for Orderly Stimulus Relations," *Proceedings of the National Academy of Sciences*, vol. 94, no. 13 (1997): 7109–7114, accessed January 2, 2024, https://doi.org/10.1073/pnas.94.13.7109; Dagmar Zeithamova, Margaret L. Schlichting, and Alison R. Preston, "The Hippocampus and Inferential Reasoning: Building Memories to Navigate Future Decisions," *Frontiers in Human Neuroscience*, vol. 6 (2012), accessed January 3, 2024, https://doi.org/10.3389/fnhum.2012.00070; Edward H. Nieh et al., "Geometry of Abstract Learned Knowledge in the Hippocampus," *Nature*, vol. 595, no. 7865 (2021): 80–84, accessed January 5, 2024, https://doi.org/10.1038/s41586-021-03652-7; André A. Fenton, "Remapping Revisited: How the Hippocampus Represents Different Spaces," *Nature Reviews Neuroscience*, vol. 25, no. 6 (2024): 428–448, accessed June 15, 2024, https://doi.org/10.1038/s41583-024-00817-x; Hristos S. Courellis et al., "Abstract Representations Emerge in Human Hippocampal Neurons During Inference," *Nature* (2024), 1–9, accessed August 10, 2024, https://doi.org/10.1038/s41586-024-07799-x.

[23] Vernon B. Mountcastle, "An Organizing Principle for Cerebral Function: The Unit Module and the Distributed System," in *The Mindful Brain: Cortical Organization and the Group-Selective Theory of Higher Brain Function* (Cambridge: The MIT Press, 1978), 7–50. Cortical columns operate similarly to hippocampal and entorhinal neural networks as well as demonstrate neuroanatomical similarities, see Jeff Hawkins and Subutai Ahmad, "Why Neurons Have Thousands of Synapses, a Theory of Sequence Memory in Neocortex," *Frontiers in Neural Circuits*, vol. 10 (2016), accessed January 11, 2024, https://doi.org/10.3389/fncir.2016.00023; Jeff Hawkins et al., "A Framework for Intelligence and Cortical Function Based on Grid Cells in the Neocortex," *Frontiers in Neural Circuits*, vol. 12 (2018): 121, accessed January 12, 2024, https://doi.org/10.3389/fncir.2018.00121.

[24] Kevin J. Mitchell, *Free Agents*, 273–275.

[25] Second-order openness means the same as being hemodynamically open; see Daniel J. Nicholson, "Reconceptualizing the Organism From Complex Machine to Flowing Stream," in *Everything*

with distributed yet constantly cooperating parallel information processing systems composed of robust, interconnected neural pathways and loops, in addition to subcortical circuits – to wit, six-layered neocortex in mammals and functionally similar structures in the avian pallium.[26]

These systems are capable of integrating and reconstructing traces of knowledge about worldly structures and the animals' internal states – affordances, motivations, goals, preferences, perceptual conclusions, and beliefs – in progressively abstract formats.[27] Thanks to constantly updating hierarchical generative models, mammals and birds can proactively undertake complex actions across diverse spatial and temporal scales. They do so to maintain the most efficient models of the world and themselves by continuously weighing the costs and benefits (consequences) of possible choices, as reinforced by interoceptive feedback.[28] This not only enables the preemptive prevention of homeostatic disturbances (detection and correction of deviations from fixed "attractive states" decoded as errors) but also the pursuit of allostatic optima – a process described within the AIN in terms of active interoceptive inference minimising expected free-energy.[29]

This brings us to the first characteristic feature of advanced agency: contrary to behaviouristic tenets, brained animals do not sit and wait for a triggering reac-

Flows: Towards a Processual Philosophy of Biology, eds. Daniel J. Nicholson and John Dupré, first edition (Process Philosophy of Biology, Oxford: Oxford University Press, 2018), 144–145.

[26] Gregory F. Ball and Jacques Balthazart, "Evolutionary Neuroscience: Are the Brains of Birds and Mammals Really so Different?," *Current Biology*, vol. 31, no. 13 (2021): R840–R842, accessed January 19, 2024, https://doi.org/10.1016/j.cub.2021.05.004.

[27] On the interconnections of neocortex with the deeper structures and its role in the control mechanisms within active inference framework, see Phan Luu, Don M. Tucker, and Karl Friston, "From Active Affordance to Active Inference: Vertical Integration of Cognition in the Cerebral Cortex Through Dual Subcortical Control Systems," *Cerebral Cortex*, vol. 34, no. 1 (2024): bhad458, accessed January 20, 2024, https://doi.org/10.1093/cercor/bhad458.

[28] This is by no means the whole story; due to lack of space, we have omitted the vital role of the microbiome and gut-brain axis; see Qinwen Wang, Qianyue Yang, and Xingyin Liu, "The Microbiota–Gut–Brain Axis and Neurodevelopmental Disorders," *Protein & Cell*, vol. 14, no. 10 (2023): 762–775, accessed March 14, 2025, https://doi.org/10.1093/procel/pwad026; Tuba Shahid Chaudhry et al., "The Impact of Microbiota on the Gut–Brain Axis: Examining the Complex Interplay and Implications," *Journal of Clinical Medicine*, vol. 12, no. 16 (2023): 1–20, accessed March 14, 2025, https://doi.org/10.3390/jcm12165231.

[29] Anil K. Seth and Karl J. Friston, "Active Interoceptive Inference and the Emotional Brain," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 371, no. 1708 (2016): 20160007, accessed June 7, 2023, https://doi.org/10.1098/rstb.2016.0007; Alexander Tschantz et al., "Simulating Homeostatic, Allostatic and Goal-Directed Forms of Interoceptive Control Using Active Inference," *Biological Psychology*, vol. 169 (2022), accessed March 14, 2025, https://doi.org/10.1016/j.biopsycho.2022.108266. For overview of the concept of allostasis in the context of the FEP see, Andrew W. Corcoran and Jakob Hohwy, "Allostasis, Interoception, and the Free Energy Principle: Feeling Our Way Forward," in *The Interoceptive Mind: From Homeostasis to Awareness*, eds. Manos Tsakiris and Helena De Preester (Oxford: Oxford University Press, 2019), 272–292.

tion stimulus – they have an intrinsic need to explore their environment selectively. Indeed, an animal with a neocortex composed of an abundance of cortical columns and extensively interconnected neural networks not only makes decisions about her conduct based on reasons – she is able to notice when reasons sometimes clash. In those internally conflicted situations, a deliberative process is activated. It is when "If… then…" capacity for counterfactual inference emerges – an ability that Donald Griffin hypothesised might be an essential brain property, given its efficiency and the survival advantage it confers.[30] The manifestations of deliberative instances (i.e., predictions comparing) have already been captured, measured, and described in detail not only in humans but also in rats. For occurrences of such indecision states, comparative neuroscientists have reintroduced Edward Tolman's term "vicarious trial and error" (VTE).[31] Then again, in light of advances in affective neuroscience – and the problematic label associated with S–R models – guided by the framework we have adopted here, we posit that this process should instead be reconceptualised as "covert epistemic inference" in service of expected free-energy minimisation.[32]

In our view, the reason philosophers used to assume this capacity to be so cognitively demanding is that during the progressing stages of evolutionary development, in primates like us with enlarged brains, it has been supplemented by multitasking abilities, language facilitating subtle variable-ratio-weighting, and collaterally, experiences of "finding oneself at crossroads." But as Griffin wrote almost three decades ago,

> Animals often select certain patterns of behavior over others of which they are quite capable, and these choices seem to be based on simple beliefs that the selected

[30] See Donald R. Griffin, "From Cognition to Consciousness," *Animal Cognition*, vol. 1, no. 1 (1998): 3–16, accessed March 21, 2024, https://doi.org/10.1007/s100710050002.

[31] VTE is a mental simulation mechanism that allows an animal to weigh the value of various options without incurring the direct costs of physically attempting them, see Adam P. Steiner and A. David Redish, "Behavioral and Neurophysiological Correlates of Regret in Rat Decision-Making on a Neuroeconomic Task," *Nature Neuroscience*, vol. 17, no. 7 (2014): 995–1002, accessed June 7, 2023, https://doi.org/10.1038/nn.3740; A. David Redish, "Vicarious Trial and Error," *Nature Reviews Neuroscience*, vol. 17, no. 3 (2016): 147–159, accessed March 5, 2023, https://doi.org/10.1038/nrn.2015.30; Brendan M. Hasz and A. David Redish, "Deliberation and Procedural Automation on a Two-Step Task for Rats," *Frontiers in Integrative Neuroscience*, vol. 12 (2018): 30, accessed March 6, 2024, https://doi.org/10.3389/fnint.2018.00030. For the relation of counterfactuals, mental-time-travel with sleep in rats, see David M. Peña-Guzmán, *When Animals Dream: The Hidden World of Animal Consciousness* (Princeton/Oxford: Princeton University Press, 2022), 88–92, 139–147. For translational research on rats and humans, see Samantha V. Abram et al., "Neural Signatures Underlying Deliberation in Human Foraging Decisions," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 19, no. 6 (2019): 1492–1508, accessed March 6, 2024, https://doi.org/10.3758/s13415-019-00733-z.

[32] See, Giovanni Pezzulo et al., "Active Inference, Epistemic Value, and Vicarious Trial and Error," *Learning & Memory*, vol. 23, no. 7 (2016): 322–338, accessed June 7, 2023, https://doi.org/10.1101/lm.041780.116.

> action will obtain some desired object or outcome. Having outgrown behaviorism there is no longer any compelling reason why ethologists should neglect the possibility that many animals choose what to do on the basis of their conscious recognition that some actions are likely to have a favorable result while others are likely to have undesirable consequences.[33]

In sum, the main strength of processual approaches to biology lies in their promise to reintegrate the concept of agency into the philosophy of animal minds and natural science. Hence, utilising the AIN conceptual toolkit profoundly bolsters our goal of deconstructing the scope of this category.[34] As a result, intentional terminology can be realistically anchored not only as a natural property of living beings but also as indicative of defining features of advanced forms of agency – with certain limits on the complexity of the unique neural patterns that the entities in question are capable of processing. We contend that genuine advanced agency commences with counterfactual inference (simulation) expressed in deliberative processes mediated by interception.[35]

Potter and Mitchell's criteria for agency, reinforced by insights from embodied Active Inference and empirical data from neuroscience, substantiate the case presented in our previous paper (*Of Rats and Men II*). Beliefs, intentions, and desires remain adequate means of description without necessarily limiting those categories to *Homo sapiens*.[36] This, in turn, enables a non-dichotomous, biocentric interpretation of discrete neuro-sensorimotor contingencies – as they are contextually enacted – in terms of a creature's (a) beliefs about her body and the world, (b) desires as an orientation towards allostatic goals – with interoceptive feedback serving as a metric for the divergence (or deviation) from expected (as preferred) states, and (c) intentions as directedness towards probabilistically possible and accessible states

[33] Griffin, "From Cognition to Consciousness," 5.

[34] For an even more processual approach to biotic agency through the lenses of FEP, see Kathryn Nave, *A Drive to Survive: The Free Energy Principle and the Meaning of Life* (London: The MIT Press, 2025).

[35] Although there is a solid case to make that counterfactual predictions originate in unconscious inferencing of raw perception as well, see Anil K. Seth, "Being a Beast Machine: The Origins of Selfhood in Control-Oriented Interoceptive Inference," in *Andy Clark and His Critics*, eds. Matteo Colombo, Elizabeth Irvine, and Mog Stapleton (New York: Oxford University Press, 2019), 247–248.

[36] It is probably not the case that folk psychology, from which intentional terms are derived, constitutes a full-fledged, albeit erroneous, theory, as Paul Churchland would have us believe (sic!). Daniel Dennett's mildly realistic intuition appears more plausible. In his arguably most pragmatically oriented paper (see Daniel C. Dennett, "Real Patterns," *Journal of Philosophy*, vol. 88, no. 1 (1991): 27–51, accessed December 24, 2024, https://doi.org/10.2307/2027085. Dennett reasoned that the predictive success of intentional concepts would be improbable without a history of consistent tracking of reliable regularities in the social domain. In line with his insight, the AIN framework seems to capture these dependencies with mathematical precision.

ZOOPHILOLOGICA.2025.15.15 p. 13/43

achievable through sequences of actions (policies). With this in mind, let us now focus on the AIN itself and other processual aspects of agency informed by this framework to identify additional features of its sophisticated forms. During this run, we will step on the vital existential question.

Meaning-Making Entangled in Action–Perception Loops

Active Inference (AIN), in its current form, has, in recent years, emerged from perception-focused predictive coding and the core tenets of the free-energy principle (FEP). Previously, within the predictive processing (PP) theory, "active inference" referred to simple sensorimotor control processes aimed at minimising proprioceptive prediction errors, and this fact sometimes leads to the misconception that it is its only application.[37] However, nowadays, this term is predominantly used in reference to models of policy and decision-making under conditions of uncertainty, directly linking perception with action and belief states with goal-directedness during the planning of a series of actions over time. Thus, AIN most closely resembles reinforcement learning and Gaussian filter models, except that it places emphasis on those aspects of cognitive processing in which the agent actively reduces uncertainty regarding present and future perception and actions – primarily through Bayesian inference, epistemic actions, and decisions based on informational observations.[38] Moreover, within the AIN, uncertainty reduction (i.e., successful prediction) is inherently rewarding, and as such, it is incorporated into the agent's generative model.

According to AIN, the brain not only anticipates incoming sensory signals and updates its world models based on them, but also takes action to ensure that these predictions come true. In other words, the brain–body influences its environment by minimising the differences between predictions and actual sensory inputs through action-taking. Actions elicit sensory observations that align with desired outcomes or goals, thereby aiding the agent in making sense of the surrounding world. Due to the recursiveness of action-perception loops, each subsequent observation of an environmental change – instigated by an action – directly

[37] Ryan Smith, Karl J. Friston, and Christopher J. Whyte, "A Step-by-Step Tutorial on Active Inference and Its Application to Empirical Data," *Journal of Mathematical Psychology*, vol. 107 (2022): 102632, accessed November 13, 2023, https://doi.org/10.1016/j.jmp.2021.102632.

[38] Mark Sprevak and Ryan Smith, "An Introduction to Predictive Processing Models of Perception and Decision-Making," *Topics in Cognitive Science*, vol. 00 (2023): 3, accessed January 2, 2024, https://doi.org/10.1111/tops.12704.

influences, though does not determine, the decisions that follow on a moment–to–moment basis.[39] The body plays a fundamental role in this agent–environment system because all modelling is mediated by it, and the overall experience emerges from the integration of spatiotemporal and phenomenal information flowing bidirectionally through it. By minimising uncertainty (which bindsevidence for the agent's model of the self and the world through sensory interactions), inferencing agents become statistical models of their eco-niches, by necessity embodying the causal structures of their *Umwelten*. Friston underscores the role of the body in an even more Ashbyan–Gibsonian, affordance-centred manner when he states that:

> [N]ot only does the agent embody the environment, but the environment embodies the agent. This is true in the sense that the physical states of the agent (its internal milieu) are part of the environment. In other words, the statistical model entailed by each agent includes a model of itself as part of that environment. This model rests upon prior expectations about how environmental states unfold over time […]. Heuristically, if I am a model of my environment and my environment includes me, then I model myself as existing.[40]

This raises a compelling question – one that is crucial both for the triptych in general and for this paper when applied to complex animals: What does it mean for an animal "to exist"? In the skilled-intentionality enactive depiction of the AIN, an animal models the relevant context-dependent possibilities for action (affordances) to approach an optimal self-generated state.[41] She maintains a predominantly ego-centred perspective on her occupied niche, and her brain–body dynamics also organise itself along this attractor. Ongoing interactions with the environment continuously reshape various aspects of her milieu according to their relevance for her (circular coupling). From this perspective, it is not the causal structure *per se* that is modelled, but rather the relationship between sensory stimulation and successfully executed actions.

These context-dependent models originate from and are tailored to the agent's previous experiences. Over time, they become increasingly context-sensitive and, concomitantly, incentivise actions that shape her surroundings to be more suitable. Thus, the structure of the generative model – besides serving to infer hidden environmental states as causes of perceptions – also guides an animal agent, through

[39] At least from the observers' vantage.

[40] Friston, "Embodied Inference," 89–90.

[41] Jelle Bruineberg, "Active Inference and the Primacy of the 'I Can'," eds. Thomas Metzinger and Wanja Wiese, *Philosophy and Predictive Processing*, vol. 5 (2017): 9, accessed August 20, 2023, https://doi.org/10.155027/9783958573062.

inter-active niche construction, towards states optimal for her.[42] This is why the generative model is not merely a model – an internal re-construction of the world, but it is also an expression of a multidimensionally intertwined agent–environment system. For the animal, it is not the most likely states of the environment that count (i.e., the "veridical" interplay of prior beliefs with current sensory evidence), but the optimal present and future states of the environment constructed with her that are, affectively speaking, of the greatest importance for this particular animal – a cognising agent.[43]

The cognising agent's decision about which action to undertake is nested in the allostatic high-level model, where interoceptive and exteroceptive (contextual) pieces of information are integrated, instantiating predictions that flow down the hierarchy and guide her autonomic adjustment or incentivise allostatic actions. It follows from PP that, in order to occupy unsurprising states, an agent's generative model hierarchy focuses on internally assessable quantities – error signals – reflecting discrepancies between what has been predicted and what has actually occurred. Given that she initially cannot foresee *a priori* which states are best for her to occupy, minimising uncertainty relative to the internal generative model of the world she embodies equates to the implicit maximisation of evidence supporting that model. In the AIN formulation, this means an agent amasses evidence for her own existence by continuously pursuing the minimisation of average prediction error (surprisal) through action, thus acting as a self-evidencing

[42] As John Dewey argued, the way the world is cognised tracks the ways it is inhabited, see Matthew Crippen and Jay Schulkin, *Mind Ecologies: Body, Brain, and World* (New York: Columbia University Press, 2020), 28.

[43] Within the PP-AIN approach, emotions are not regarded as ontologically separable from cognitive processes or indicative of a special type of cognition. Instead, they form an aspect of the generative model that integrates a predictive processing system. Like other categories of experience, emotions are compressions (*ad hoc* concepts) modulated by the core affect, with dimensions of arousal and valence permeating states of beliefs and desires. Therefore, the predictive mind is inherently emotional. See Kathryn Nave et al., "Wilding the Predictive Brain," *WIREs Cognitive Science*, vol. 11, no. 6 (2020): e1542, accessed December 20, 2024, https://doi.org/10.1002/wcs.1542. Due to lack of space, we have to omit the details of the affective dimension of the AIN framework (for the role of emotions in cognitive processes, see Mark Miller and Andy Clark, "Happily Entangled: Prediction, Emotion, and the Embodied Mind," *Synthese*, vol. 195, no. 6 (2018): 2559–2575, accessed August 25, 2023, https://doi.org/10.1007/s11229-017-1399-7; for Pankseppian approach to emotions within the FEP framework see, Mark Solms, *The Hidden Spring: A Journey to the Source of Consciousness* (New York: W. W. Norton & Company, 2021); Mark Solms and Karl Friston, "How and Why Consciousness Arises: Some Considerations from Physics and Physiology," *Journal of Consciousness Studies*, vol. 25 (2018): 202–238, accessed August 25, 2023, https://discovery.ucl.ac.uk/id/eprint/10057681/1/Friston_Paper.pdf?ref=quillette.com; for view conflicting with basic emotion theory, yet currently under the same theoretical umbrella (FEP), see Lisa Feldman Barrett, "The Theory of Constructed Emotion: An Active Inference Account of Interoception and Categorization," *Social Cognitive and Affective Neuroscience*, vol. 12, no. 1 (2017): 1–23, accessed August 4, 2023, https://doi.org/10.1093/scan/nsw154.

entity.[44] Therefore, the AIN answer to the question of what it means for an animal agent to exist is circular – to exist is to act in such a way as to be one's own proof of existence. This answer may seem absurd *prima facie*, but it is firmly grounded in the concept of abduction, first developed by Charles Sanders Peirce. Abductive inference is a notion closely related to the idea of inference to the best explanation, which in this context, is more about the likelihood of expected outcomes.[45] As Jacob Hohwy writes:

> In an inference to the best explanation, some evidence is explained by some hypothesis – that best explains the evidence – which is then inferred as true or probable. Often, the evidence for the hypothesis is just the very evidence it explains. This creates a benign circle where the hypothesis explains the evidence, and the evidence is evidence for the hypothesis. The explanatory power of the hypothesis ensures the hypothesis has evidence for itself – it is self-evidencing […]. There is no outside help or evidence to establish the hypothesis (conceived as the system's internal generative model) – the only evidence is the evidence the model can account for. The system needs to configure or organise its own internal states (conceived as the model's parameters and states) in order to make the evidence (or sensory input) as expected or unsurprising as possible. As it accomplishes this task, it accumulates evidence for itself.[46]

The notion of self-evidencing constitutes another disruption of classical perspectives – this time regarding the relationship between perception, action, and epistemology.[47] AIN's central idea is that perception and action simultaneously con-

[44] Sounds sophisticated, but it means no more than that an animal's brain is constantly guessing; see Jakob Hohwy, "Conscious Self-Evidencing," *Review of Philosophy and Psychology*, vol. 13, no. 4 (2022): 809–828, accessed August 27, 2023, https://doi.org/10.1007/s13164-021-00578-x; Anil K. Seth, "Inference to the Best Prediction," in *Open MIND*, eds. Thomas Metzinger and Jennifer Windt (Frankfurt am Main: MIND Group, 2015).

[45] When invoking the idea of abductive inference, Hohwy has somehow forgotten its originator. For discussions about animal abduction, see Lorenzo Magnani, "Animal Abduction," in *Abductive Cognition: The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning*, ed. Lorenzo Magnani, Cognitive Systems Monographs (Berlin–Heidelberg: Springer, 2009), 265–316; Mariana Vitti Rodrigues and Claus Emmeche, "Abduction: Can Non-Human Animals Make Discoveries?," *Biosemiotics*, vol. 10 (2017): 295–313, accessed August 26, 2023, https://doi.org/10.1007/s12304-017-9300-0. See also relevant in *Of Rats and Men II*, footnote 35 on the linkage of perceptual inference with Peircean semiotics embedded in synechism.

[46] Hohwy, "Conscious Self-Evidencing," 8.

[47] This concept deprives the Cartesian model of its departure point. Friston has summarised it as follows: "I will only exist iff (sic!) I am a veridical model of my environment. Put even more simply; 'I think therefore I am, iff I am what I think.' This tautology is at the heart of the free-energy principle and celebrates the circular causality that underpins much of embodied cognition." See Friston, "Embodied Inference," 90.

tribute to minimising discrepancies between actual and predicted sensory inputs. However, departing from the rationalistic (Helmholtzian) depiction of cognition, the goal is not merely to infer hidden causes from the environment but rather to steer agent–environment interactions towards outcomes that are optimal for the agent's flourishing. In line with pragmatic readings of evolutionary theory, the ecological-enactive approach posits that cognisers are not in pursuit of recognising objective reality; instead, they aim to maximise the evidence substantiating their predictions.[48]

In other words, agents engage in actions that affirm their generative models, thereby providing evidence for their own existence through the changes they effect in the world. In this sense, for an advanced rendition of cortically equipped social animals, the agent must distinguish between her perceptual and interoceptive sensations resulting from her own actions and those caused by external factors (e.g., other agents or events). This process can occur on many levels of organisation through various means, though it originates at the level of the pre-reflective embodied self – an inchoate experience of being an active body – that gives rise to the impression of authorship over the action performed.[49]

How does the content discussed so far integrate into the task of reconstructing advanced agency using a minimally sophisticated model, namely rats? The capacity for causal inferencing in these creatures, including their ability to distinguish their own actions from those they observe, remains a subject of considerable controversy (albeit primarily among behaviourists). Nonetheless, it has been extensively explored over the past two decades by leading comparative neuroscientists.[50] Based

[48] It follows that the initial Helmholtzian metaphor of the brain as a scientist must account for the fact that this scientist is "crooked" because she tends to bend her hypothesis to her "optimistic generative model" all the time.

[49] See Anil Seth, "Interoceptive Inference, Emotion, and the Embodied Self," *Trends in Cognitive Sciences*, vol. 17, no. 11 (2013): 565–573, accessed August 28, 2023, https://doi.org/10.1016/j.tics.2013.09.007; Jakub Limanowski and Felix Blankenburg, "Minimal Self-Models and the Free Energy Principle," *Frontiers in Human Neuroscience*, vol. 7 (2013), accessed August 23, 2023, https://www.frontiersin.org/articles/10.3389/fnhum.2013.00547; Bruineberg, "Active Inference and the Primacy of the 'I Can'," 12.

[50] See Kenneth J. Leising et al., "The Special Status of Actions in Causal Reasoning in Rats," *Journal of Experimental Psychology: General*, vol. 137, no. 3 (2008): 514–527, accessed December 10, 2023, https://doi.org/10.1037/0096-3445.137.3.514; Aaron P. Blaisdell et al., "Causal Reasoning in Rats," *Science*, vol. 311, no. 5763 (2006): 1020–1022, accessed December 12, 2023, https://doi.org/10.1126/science.1121872; Kosuke Sawa, "Predictive Behavior and Causal Learning in Animals and Humans," *Japanese Psychological Research*, vol. 51, no. 3 (2009): 222–233, accessed December 16, 2023, https://doi.org/10.1111/j.1468-5884.2009.00396.x; Michael R. Waldmann et al., "Rats Distinguish Between Absence of Events and Lack of Evidence in Contingency Learning," *Animal Cognition*, vol. 15, no. 5 (2012): 979–990, accessed December 17, 2023, https://doi.org/10.1007/s10071-012-0524-8; for deflationary associative account see Dominic M. Dwyer, Judy Starns, and Robert C. Honey, "'Causal

on the discussion thus far, it seems to naturally follow the functional organisation of advanced cognitive systems, serving as a facilitator of predictive precision (inverse variance) for density of probability distributions (expected states). However, the rabbit hole of self-evidencing extends deeper. Before progressing to the next section of this article, it is necessary to allocate some space to explore this topic further.

Sophistically Self-Evidencing Rat?

A key aspect of advanced self-evidencing is active counterfactual inference – previously mentioned – which need not be conscious, much less discursive.[51] This kind of constructivist simulation is carried out continuously by animal agents.[52] To perform it, they must maintain internal models of the world's causal structures, coupled with models of themselves within those structures (self-models).[53] They continually infer the hidden causes of their sensory input, informed by probabilistic

Reasoning' in Rats: A Reappraisal," *Journal of Experimental Psychology: Animal Behavior Processes*, vol. 35, no. 4 (2009): 578–586, accessed December 17, 2023, https://doi.org/10.1037/a0015007. For a compromise approach, see Dominic M. Dwyer and Michael R. Waldmann, "Beyond the Information (Not) Given: Representations of Stimulus Absence in Rats (Rattus Norvegicus)," *Journal of Comparative Psychology*, vol. 130, no. 3 (2016): 192–204, accessed December 17, 2023, https://doi.org/10.1037/a0039733. For an overview, see Francisco J. Silva and Kathleen M. Silva, "Causal Reasoning in Rats (Blaisdell et al. 2006)," in *Encyclopedia of Evolutionary Psychological Science*, eds. Todd K. Shackelford and Viviana A. Weekes-Shackelford (Cham: Springer International Publishing, 2021), 936–940.

[51] Language is neither a necessary nor sufficient condition for thinking; see Evelina Fedorenko, Steven T. Piantadosi, and Edward A. F. Gibson, "Language Is Primarily a Tool for Communication Rather than Thought," *Nature*, vol. 630, no. 8017 (2024): 575–586, accessed July 2, 2024, https://doi.org/10.1038/s41586-024-07522-w.

[52] The notion that counterfactual inference does not have to be discursive is evidenced not only by examples of individuals who base their inferential processes (thinking) on visual imagination (e.g., Temple Grindin) but also by the recently uncovered (and not as rare as one might assume) cases of people who despite the absence of inner speech (Anauralia, Anendophasia) otherwise function completely normally. See also surprising experimental results predating this discovery in Russell T. Hurlburt and Sarah A. Akhter, "Unsymbolized Thinking," *Consciousness and Cognition*, vol. 17, no. 4 (2008): 1364–1374, accessed December 15, 2023, https://doi.org/10.1016/j.concog.2008.03.021. For Anauralia (Anendophasia) research, see Rish P. Hinwar and Anthony J. Lambert, "Anauralia: The Silent Mind and Its Association With Aphantasia," *Frontiers in Psychology*, vol. 12 (2021), accessed December 20, 2023, https://doi.org/10.3389/fpsyg.2021.744213; Johanne S. K. Nedergaard and Gary Lupyan, "Not Everybody Has an Inner Voice: Behavioral Consequences of Anendophasia," *Psychological Science* , vol. 35(7) (2024), accessed May 10, 2024, 09567976241243004, https://doi.org/10.1177/09567976241243004.

[53] "The brain is the statistical model of the world it inhabits," including the agent's body model. That way, self-awareness is encrypted in the brain-function as a default.

feedback derived from statistical variables sampled from the generative process (the world). This process forms the interpretative basis for computations within their generative models.[54] In doing so, they reduce discrepancies between their internal generative models and the actual states – *via* the refinement of sensory information adjusted to spatiotemporal error feedback. In consequence, such animals are essentially future-oriented, able to flexibly minimise expected uncertainty (free energy) over various timespans. As it turns out – prediction and inference "are two sides of the same coin."[55]

As we sketched above, to remain within the boundaries of viable states and flourish, an animal uses priors – beliefs about the estimated conditions she occupies as viability measures (*vel* probability distributions). Priors (Bayesian beliefs) represent expectations about the states an animal occupies at any given moment and about future states; they are partially acquired through experience (learned) and partially shaped during the evolutionary process.[56] Advanced organisms regulate their homeodynamics by adjusting (in PP – through adaptation or perceptual inference – "changing mind") or by pursuing additional confirmation for those priors, that are challenged by sensory feedback (in AIN – through "changing the world"). External states mediate changes in perception and must be accounted for. However, if the precision of a given prediction is sufficiently high (corresponding with the degree of confidence), it is the external states that must conform. This iterative back-and-forth is grounded in perception-action loops, exerting relative control over worldly states across multiple timescales.

Yet, not every possible action is worth performing; hence, internal models constantly generate competing predictions, modelling each in accordance with possible outcomes – that is, as counterfactuals. In animals lacking the language network,[57] active counterfactual inferencing (covert epistemic inference) is processed

[54] The generative model and the generative process represent two different ways of generating sensory data. The generative process refers to the actual structure of the world, that is, to how sensory data are generated. In contrast, the generative model is a construct used by the agent to draw inferences about the latent causes behind sampled data. These two modes of information flow can differ substantially.

[55] Bruineberg, "Active Inference and the Primacy of the 'I Can'," 4.

[56] It is also in line with the pragmatists' view, by the fact that since the Metaphysical Club (with the exception of neopragmatists engaged in the linguistic turn), they saw belief less as a propositional attitude and more as "a habit of expectation which is conditioned by the force of experience," see Cheryl Misak, "The Metaphysical Club," in *The Routledge Companion to Pragmatism*, eds. Scott Aikin and Robert Talisse (New York–London: Routledge, 2023), 7, or in Alexander Bain's version, as "a representation poised to guide an animal's voluntary actions or exertions," Aaron Zimmerman, "Pragmatism in the Philosophy of Mind," in *The Routledge Companion to Pragmatism*, 228.

[57] See Evelina Fedorenko, Anna A. Ivanova, and Tamar I. Regev, "The Language Network as a Natural Kind Within the Broader Landscape of the Human Brain," *Nature Reviews. Neuroscience*, vol. 25, no. 5 (2024): 289–312, accessed June 8, 2024, https://doi.org/10.1038/s41583-024-00802-4. See also footnotes 50, and 51.

through neuronal structures responsible for the previously mentioned VTE (see footnotes 30 and 31). In an AIN depiction, this type of internally driven process represents a computational race between competing simulations of possible courses of action (policies), with the victor of this rivalry being the option that promises the greatest reduction in uncertainty. To achieve this, the cognitive system relies on an internal model of the environment and self. The more experience a creature accumulates through interacting with the world, the more detailed and trustworthy these models become. This, in turn, leads to increased flexibility and optimistic bias in action selection (or "willpower"). This mechanism precisely accounts for the surprising driving skills exhibited by the rats in the experiment we referenced in our first paper of the triptych.[58] Rats group-raised in affordances-rich habitats hesitated less before entering and mastering the vehicle and enjoyed driving for its own sake, without external rewards. This behaviour sharply contrasts with the lack of competence and "willpower" displayed by conspecifics reared in pairs in standard "boring" cages. Moreover, skilled drivers experienced higher stress levels when forced to occupy "the passenger seat," demonstrating the sense of agency (SoA) at work.[59] We will return to this concept later in the paper.

Though the applications of the AIN framework to empirical investigations remains at relatively early stages of development, one can already point to promising results supporting its underlying assumptions. The fertility of predictive frames is demonstrated by a series of recent findings involving rodents, published in the journal *Nature*. For example, the use of novel techniques – such as virtual environments and optogenetics – has confirmed both predictive mechanics and the existence of predictive maps in rodent brains.[60] Adding to this, researchers have found that mice can recognise themselves in a mirror when prompted by tactile cues (the spot painted on the creatures' heads needs to be heavy enough to draw their attention), provided they have had social experiences with conspecifics of the same strain.[61] A parsimonious hypothesis would suggest that if mice can interpret

[58] Słuszkiewicz, "Of Rats and Men I," 15–16.

[59] See Elizabeth Crawford et al., "Enriched Environment Exposure Accelerates Rodent Driving Skills," *Behavioural Brain Research*, vol. 378 (2020): 112309, accessed June 10, 2023, https://doi.org/10.1016/j.bbr.2019.112309.

[60] To list a few: Nicholas Cole et al., "Prediction-Error Signals in Anterior Cingulate Cortex Drive Task-Switching," *Nature Communications*, vol. 15, no. 1 (2024): 7088, accessed August 27, 2024, https://doi.org/10.1038/s41467-024-51368-9; David G. Lee et al., "Perirhinal Cortex Learns a Predictive Map of the Task Environment," *Nature Communications*, vol. 15, no. 1 (2024): 5544, accessed August 28, 2024, https://doi.org/10.1038/s41467-024-47365-7; Shohei Furutachi et al., "Cooperative Thalamocortical Circuit Mechanism for Sensory Prediction Errors," *Nature*, no. 633 (2024): 1–9, accessed August 28, 2023, https://doi.org/10.1038/s41586-024-07851-w.

[61] Jun Yokose, William D. Marks, and Takashi Kitamura, "Visuotactile Integration Facilitates Mirror-Induced Self-Directed Behavior through Activation of Hippocampal Neuronal Ensembles in

perceptual, interoceptive, and proprioceptive information to construct a model of themselves (i.e., they self-reference – develop an interoceptive model of the inferred causes of their own sensorium), then rats might be even more capable of such self-modelling. However, this conjecture has not yet been experimentally confirmed. Regardless, these experimental findings support the thesis that rodents can form extensive models of the world and, at the very least, rudimentary cognitive models of themselves – two fundamental features needed for the emergence of self-evidencing on the level of psychophysics.

According to Hohwy, humans, and presumably some other animals, engage in even more profound forms of counterfactual computational operations, which he terms "sophisticated counterfactual inference."[62] A creature capable of such inferencing may plan in multiple steps, construct counterfactual models of actions and events, and confront them with one another, thus separating representations from direct experience. Such a creature can evaluate actions based on both implicit and explicit models of beliefs about herself and the world, even when confronting imminent information that contradicts these models, thereby gaining temporal depth.[63]

Although demonstrating this type of inference in rats is considerably more challenging, experimental evidence suggests that they can engage in it. For instance, researchers have used state-of-the-art techniques to introduce rats to virtual reality *via* a brain-machine interface and have provided rodents with visual data (representations induced directly in the hippocampal place cells, whose activation could be read back). In this way, rats have acquired novel skills for locomotion within the virtual environment. They soon abandoned physical limb movement, instead focusing on conscious, voluntary control over the exploration of the environment in a goal-oriented manner. Using only their imagination, they were even able to remotely translocate virtual objects (akin to the feats of Jedi Knights), exhibiting the capacity to maintain and mentally control the movement of detached-from-world virtual representations.[64] This experiment not only demonstrated the purposive mental behaviour of rodents but, within the AIN framework, can also be interpreted as evidence of the ability of these animals to update their model of the world and themselves in the face of contradictory sensory information. In this context, updating the model has led to the abandonment of such an essential behaviour as overt motor execution (utilisation of motor priors), as it was unnecessary given the circumstances.

Mice," *Neuron*, vol. 112, no. 2 (2024): 306–318.e8, accessed June 20, 2024, https://doi.org/10.1016/j.neuron.2023.10.022.

62 Hohwy, "Conscious Self-Evidencing," 9.

63 Hohwy, "Conscious Self-Evidencing," 9.

64 Chongxi Lai et al., "Volitional Activation of Remote Place Representations with a Hippocampal Brain–Machine Interface," *Science*, vol. 382, no. 6670 (2023): 566–573, accessed December 15, 2023, https://doi.org/10.1126/science.adh5206.

Moving further, Hohwy surmises that agents capable of sophisticated counterfactual inference should exhibit analogous responses to perceptual illusions.[65] In addition, he argues that sophisticated counterfactual inference – which utilises action not only to achieve expected (preferred) states but also to reduce uncertainty for its own epistemic value – indicates perceptual consciousness. This capability enables an explorative *modi operandi* that balances potential reward-oriented actions for epistemic value with those aimed at utility (i.e., direct reward-seeking).[66] Hohwy predicts that if an agent lacks sufficient information about the distribution of rewards in the environment or knowledge about the possibilities of obtaining them (which is predominantly the case outside the lab), she will try to explore more. She will decide to act for a reward only after reaching a certain threshold of certainty. This type of behaviour implies that the agent has a representation of her own uncertainty (recognition that she does not know something – a mark of metacognition). This mental skill would have significant adaptive value in a volatile social environment, where states of uncertainty (i.e., information overload or noise) occur frequently. Therefore, the ability to perform sophisticated counterfactual inferences may be an essential property of advanced consciousness. In short, agents capable of sophisticated self-evidencing – that is, approaching uncertainty reduction flexibly through selective sampling of the environment – must effectively manage their urges, that is, exhibit self-control (or gratification delay).

The most recent experimental findings support the criteria outlined above:

[1] Rats can solve problems in both model-based and model-free variants – the former being constitutive for generative model formation.[67] Although it would be more precise to say that rats form complex internal models – since such a capacity is required to display physiological reactions to perceptual illusions – this would

65 Hohwy, "Conscious Self-Evidencing," 10.

66 Ibid.

67 See, for example, Hasz and Redish, "Deliberation and Procedural Automation on a Two-Step Task for Rats." This capacity has also been confirmed in the virtual reality experiment described earlier and in studies mentioned in footnotes 30, and 49. We should note that, unlike Reinforcement Learning (RL) paradigm, AIN does not recognise model-free – model-based dichotomy. RL focuses on optimising behavior on reward and punishment signals, often without direct reference to Bayesian inference about sensory causes. In the model-free approach, the agent forms policies directly from experience by learning the stimulus-response mapping without building an internal model of the environment. AIN is a belief-based approach. According to it, the reward function is part of the probabilistic generative model. It is interpreted in the context of preferred states that the agent expects and strives for by minimising expected free energy. Therefore, the value function is not a signal external to the agent but a reflection of its internal beliefs (probability distributions) and goals. "Model-free" systems are simply systems that form less complex models adapted to certain properties of their eco-niche. See Thomas Parr, Giovanni Pezzulo, and Karl J. Friston, *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior* (New York: The MIT Press, 2022), 203, 206–209, 269.

not be feasible without the propensity of these creatures to construct detailed top–down perceptual models based on sensory predictions.[68]

[2] In a recently designed experiment examining the learning process of abstract-object categorisation based on the trade-off between speed and accuracy, it was observed that these animals optimise the learning by initially forgoing rewards in favour of spending time acquiring more information.[69] In the AIN depiction, this would suggest that rats, at least sometimes, strategically prioritise epistemic value over utility (i.e., they prefer epistemic policy over immediate rewards or pragmatic value), precisely as Hohwy anticipated.

[3] Moreover, rats are capable of monitoring the accuracy of their decision-making processes. Researchers have established this higher-ordered metacognitive capacity by designing conditions in which rats behaviourally "report" the experienced prediction error of their previous decisions concerning time-interval categorisation. This experiment is the first to demonstrate the metacognitive efficacy of self-monitoring in rodents beyond trial-and-error training through simple "uncertainty response" or "confidence judgment" tasks. These animals show extraordinary precision in retrospectively assessing their own actions, as evidenced by their independent generation of time intervals in single trials. According to the researchers, apart from assessing the accuracy of internal representations of the world and estimating discrepancies without external feedback, rats also represent both the average value and the uncertainty of the internal time perception.[70]

[4] Lastly, there is also evidence that rats are capable of exhibiting substantial self-control (self-supervision) while simultaneously engaging in counterfactual inferencing.[71] We will return to these findings after connecting counterfactual self-evidencing with the research on the experience of action ownership.

[68] Dmitrii Vasilev, Isabel Raposo, and Nelson K. Totah, "Brightness Illusions Evoke Pupil Constriction Preceded by a Primary Visual Cortex Response in Rats," *Cerebral Cortex*, vol. 33, no. 12 (2023): 7952–7959, accessed December 9, 2023, https://doi.org/10.1093/cercor/bhad090.

[69] Javier Masís et al., "Strategically Managing Learning During Perceptual Decision Making," eds. Valentin Wyart, Michael J. Frank, and Konstantinos Tsetsos, *eLife*, vol. 12 (2023): e64978, accessed December 20, 2023, https://doi.org/10.7554/eLife.64978.

[70] Tadeusz W. Kononowicz, Virginie van Wassenhove, and Valérie Doyère, "Rodents Monitor Their Error in Self-Generated Duration on a Single Trial Basis," *Proceedings of the National Academy of Sciences*, vol. 119, no. 9 (2022): e2108850119, accessed December 14, 2023, https://doi.org/10.1073/pnas.2108850119.

[71] Vincent Laurent and Bernard W. Balleine, "Factual and Counterfactual Action-Outcome Mappings Control Choice Between Goal-Directed Actions in Rats," *Current Biology*, vol. 25, no. 8 (2015): 1074–1079, accessed December 16, 2023, https://doi.org/10.1016/j.cub.2015.02.044; Steiner and Redish, "Behavioral and Neurophysiological Correlates of Regret in Rat Decision-Making on a Neuroeconomic Task."

The concept of self-evidencing converges not only with biological, psychophysical, and psychophysiological dimensions but also – as we have prompted earlier – with research on the Sense of Agency (SoA), particularly with its most prominent model, the comparator model. In philosophical terms, SoA can be defined as an experience of being in control of one's actions and, through that, of influencing worldly events.[72] Within the AIN framework, SoA can be divided into two parts: the bodily feeling of agency (BFoA) and the judgment of agency (JoA). The former is a pre-reflective feeling accompanying voluntary action – a sense of being in control – while the latter is the ability to refer to oneself as an author of a given action. Both these experiences can be described as parallel generative models of the world, underpinning processual (abductive) inferences to the best explanation (from the agent's perspective) regarding the likely causes of the current and future states of the agent. In this conceptualisation, while egocentric predictions produce BFoA, allocentric give rise to JoA.[73]

According to the comparative model, an agent can identify bodily ownership of her actions by predicting sensory impressions resulting from self-initiated movement and comparing them with actual sensory feedback. The comparator model is based on a cybernetic principle, proposing that agents maintain their autonomy by suppressing and responding to (controlling for) environmental perturbations. The brain constantly predicts the outcomes of an agent's actions and compares the efference copies of motor commands with incoming sensory feedback. This process leads to updates of the predictions or modifications of subsequent actions. In this way, the comparator model fits within the AIN framework, but it gets simpler as AIN negates the necessity for duplicating the motor command.[74]

The comparator model is used to explain SoA through the psychophysical measurement of the "intentional binding," an implicit effect considered a hallmark of volition.[75] It is a perceptual illusion in which the individuals perceive their volun-

[72] Shaun Gallagher, "Philosophical Conceptions of the Self: Implications for Cognitive Science," *Trends in Cognitive Sciences*, vol. 4, no. 1 (2000): 14–21, accessed August 28, 2023, https://doi.org/10.1016/S1364-6613(99)01417-5.

[73] Pantelis Leptourgos and Philip R. Corlett, "Embodied Predictions, Agency, and Psychosis," *Frontiers in Big Data*, vol. 3 (2020), accessed August 18, 2023, https://www.frontiersin.org/articles/10.3389/fdata.2020.00027.

[74] Andy Clark, *The Experience Machine*, 79.

[75] Patrick Haggard, "Sense of Agency in the Human Brain," *Nature Reviews Neuroscience*, vol. 18, no. 4 (2017): 196–207, accessed August 29, 2023, https://doi.org/10.1038/nrn.2017.14.

tary actions and subsequent effects as contracted in time compared to actions they did not initiate.[76] Various findings regarding this phenomenon have been reported since Patrick Haggard and colleagues discovered it in the context of empirical inquiry into foundational aspects of the experience of free will two decades ago.[77] Until now, the entire paradigm has been studied on humans, because such experiments require verbal cooperation. In the case of complex animal cognition and AIN, an interesting question would be to ask how is the space of possible hypothesis "crooked" in such a way as to enable an animal agent bend *Umwelt* to optimal conditions for her.[78] The intentional binding effect would be an ideal candidate for demonstrating the intimate coupling of an agent with her environment. If the contraction of perceived time between action and outcome (along the dopamine fluctuations) coupled with minimal prediction-error (a successful prediction) were observed, it could be interpreted as a form of interoceptive pre-narrative self-evidencing – the sense of ownership (or "mineness") of the performed action.

As it turns out, there is a potential method to test this question, thanks to an experimental setting put forward by Antonella Tramacere and Colin Allen.[79] The authors have initially aimed to restructure the debate on causal cognition in animals. They proposed adopting the "intentional binding" paradigm by translating measurements of the subjective timing relationship between event outcomes and the SoA to investigate implicit causal inference in animals. Tramacere and Allen suggest that their experimental design can serve as a test for the existence of a non-linguistic level of representations necessary to explain some of the animal problem-solving

[76] We use the term "volition" here in a purely neurobiological sense as the neural and cognitive machinery involved in decision-making and choice. Volition as a component of executive function is not exclusive to humans; it encompasses many aspects of action – from learning through experience, extracting information from memory to preferences development and planning. Nowadays, it has well-identified neural structures, and there is nothing mysterious about it; see Philip Ball, *The Book of Minds: Understanding Ourselves and Other Beings, from Animals to Aliens* (London: Picador, 2023), 415–423.

[77] Patrick Haggard, Sam Clark, and Jeri Kalogeras, "Voluntary Action and Conscious Awareness," *Nature Neuroscience*, vol. 5, no. 4 (2002): 382–385, accessed August 29, 2023, https://doi.org/10.1038/nn827.

[78] Bruineberg, "Active Inference and the Primacy of the 'I Can'," 9–12. More about von Uexküll's concept of *umwelt* see Colin Allen, "*Umwelt* or *Umwelten*? How Should Shared Representation Be Understood Given Such Diversity?," *Semiotica*, no. 198 (2014): 137–158, accessed August 23, 2023, https://doi.org/10.1515/sem-2013-0105. About the relation of *umwelt* with Peirce's broadened view on logic, see Kalevi Kull, Dmitry Kozhevnikov, and Magdalena Kozhevnikova, "O logice zwierzęcych umweltów. Subiektywna teraźniejszość zwierząt, czyli zoosemiotyka wyboru i uczenia się," *Zoophilologica. Polish Journal of Animal Studies*, vol. 11, no. 1 (2023): 1–20, accessed August 30, 2023, https://doi.org/10.31261/ZOOPHILOLOGICA.2023.11.06.

[79] Antonella Tramacere and Colin Allen, "Temporal Binding: Digging Into Animal Minds Through Time Perception," *Synthese*, vol. 200, no. 1 (2022): 1–24, accessed June 1, 2023, https://doi.org/10.1007/s11229-022-03456-w.

skills. They opted for a multilevel framework of mental representations in which an intermedial level is accessible to at least some animals and prelinguistic children.[80] To advance their proposition, they adopt a vital feature of the intentional binding effect called "temporal binding," picking up the finding that if a human agent recognises her particular action as causing an effect, she experiences an above-mentioned illusionary compression of time – even if the actual cause is due to external factors (e.g., another agent or event). In an experimental design called "synchrony judgment task," the researchers use the rat solely as an exemplary model. It is worth noting, however, that when one considers the cumulative results of experimental studies indicating the extensive cognitive-affective abilities of this species,[81] rats appear to be a near-perfect prototype for research in non-languaging mammals.

In short, the basic idea is as follows: Rats are trained to distinguish time intervals between two signals (T1, T2). Depending on the interval length, they then choose food rewards (e.g., sugar pellets) at designated spots in a maze, 'S' for short intervals and 'L' for extended intervals. Next, rats are divided into pairs ("actor" and "observer"). The actor can choose an action (pressing the lever, button, or key), therewith evoking signal T1, and estimates the time interval between the initiated action and the following signal T2. The observer, by contrast, plays a witness in a similar maze without control over T1. The results are compared: if the experimental "actors" group significantly favours the 'S' site over the "observers" group, this would suggest the presence of a temporal binding effect (and, by egocentric priority, intentional binding). If no difference is observed, then "there is no there there." Should the effect's presence be confirmed, a plethora of inquiry avenues might open for researchers.[82]

Yet, in the paper, we pursue a slightly divergent agenda. At this point, we refer back to our discussion of Hohwy's criteria for sophisticated self-evidencing to reconnect sophisticated counterfactual inference with the SoA and its subjective manifestation in rats as advanced cognitive agents. If our claim holds merit, we can predict

[80] We suggest that grounded cognition theory would be helpful in this project; see Lawrence Barsalou, "Grounded Cognition," *Annual Review of Psychology*, vol. 59 (2008): 617–645, accessed August 30, 2023, https://doi.org/10.1146/annurev.psych.59.103006.093639; Lisa Feldman Barrett, "The Theory of Constructed Emotion: An Active Inference Account of Interoception and Categorization."

[81] For a review and interpretation of rats' cognitive and affective competences in the context of embodied morality research, see Zbigniew Słuszkiewicz, "Szczur w badaniach psychologii społecznej i w naukach kognitywnych. *Bestia Sacer* czy istota motywowana moralnie?," *Zoophilologica. Polish Journal of Animal Studies*, vol. 10, no. 2 (2022): 1–70, accessed August 30, 2023, https://www.journals.us.edu.pl/index.php/ZOOPHILOLOGICA/article/view/11748. For current overview (as a background for ethical argument), see Inbal Ben-Ami Bartal, "The Complex Affective and Cognitive Capacities of Rats," *Science*, vol. 385, no. 6715 (2024): 1298–1305, accessed October 10, 2024, https://doi.org/10.1126/science.adq6217.

[82] See the "Discussion" section in Tramacere and Allen, "Temporal Binding."

that successful implementation of the "temporal binding" experiment in a more elaborate form of this experiment, incorporating the summation of the temporal average across a sequence of actions, will demonstrate that rats can weigh between possible policies based on the temporal duration of the previously experienced events. In doing so, they will tend to select scenarios that entail the least ambiguity and highest control (i.e. the lowest expected free energy). In this context, a lack of ambiguity means that the animal can map predictions of actions onto their outcomes as faithfully as possible, effectively forecasting the future states of the world before executing the chosen sequence of actions. The more precise the prediction – meaning, the less expected free energy (or surprise) the animal's generative model augurs before implementation – the greater the likelihood that it will be selected.

According to Hohwy, it is the anticipated precision of the predicted results – the consistency of expectations with observations – that determines the policy selection, whereas precision optimisation can be captured in terms of fine-tuning the attentional mechanism.[83] In this scenario, the agent operates on the principle of a self-fulfilling prophecy, as if she believed that she had performed actions whose effects were optimistically predicted.[84] Since the selected policy pre-empts prediction errors, if no disruptions occur during the alignment of predictions with sensory data (perceptive, proprioceptive, and interoceptive), the process will proceed smoothly, generating minimal error signals at the lower levels of processing. Instead, the agent's reward system will be flooded with dopamine, which heralds prediction success as an equivalent to the expected reward both in humans and rats.[85]

Given what has been argued for in our previous and current paper, and what is of particular interest to us, the interpretation of rodents' actions during the experiment in question by means of the FEP's toolkit could provide further support for the traditional framework of folk psychology – without engendering redundant controversies. The arguments presented in *Of Rats and Men II* regarding intentional nomenclature in the context of PP become even more evident within the AIN. Their conceptual schema renders the language of agency in the animal context – regarded hitherto by conservative scholars as akin to heresy – much more straightforward; our laboratory rats do not press the lever, button, or key because they unconsciously associate it with sugar pellets. Rather, they desire the pellets and, based on experi-

[83] See Hohwy, "Conscious Self-Evidencing," 14–15.

[84] Elizabeth L. Fisher and Jakob Hohwy, "The Universal Optimism of the Self-Evidencing Mind," *Entropy*, vol. 26, no. 6 (2024): 518, accessed June 21, 2024, https://doi.org/10.3390/e26060518.

[85] Wolfram Schultz, Peter Dayan, and Pendleton Read Montague, "A Neural Substrate of Prediction and Reward," *Science*, vol. 275, no. 5306 (1997): 1593–1599, https://doi.org/10.1126/science.275.5306.1593, accessed November 16, 2023, Wolfram Schultz, "Dopamine Reward Prediction Error Coding," *Dialogues in Clinical Neuroscience*, vol. 18, no. 1 (2016): 23–32, accessed November 17, 2023, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4826767/.

enced approximation of statistical regularities, believe they have predicted the likelihood of obtaining them as a result of the actions they intended.

If rats experience shortened time intervals following action, it implies they implicitly infer their own authorship, thereby reassuring (self-evidencing) their self-models. We surmise that when prediction error feedback occurs minuscule, an influx of dopamine into the reward accompanying the prediction of the result of action (model evidence) evokes an experience of contracted time between lever pressing and the outcome.[86] This constitutes the basis of the sense of authorship as reported behaviourally by an animal running to the 'S' spot in the maze. Such behavioural – but also psychophysical and psychophysiological – expressions of the sense of ownership of performed actions could, *per analogiam,* offer us a glimpse into the nonhuman self-evidencing, or, to paraphrase Thomas Nagel's famous gap-pointing, into "how it is like to be an acting rat."[87] Certainly, demonstrating temporal binding in rats does not resolve the Chalmersian "hard problem of consciousness," but it can be seen as an advancement on the "real problem" advocated by Anil Seth. Simultaneously, it integrates this aspect of agency's subjective experience in both humans and other species without becoming mired in the "qualia debates."[88]

Sophisticated Agents Within the Cognitive Light Cone

What, then, does rodents' self-evidencing have to do with FW besides their self-rewarding, dopamine-fluctuation-based spontaneity during voluntary exploration – which has already been demonstrated?[89] According to Karl Friston,[90] within

[86] For a detailed overview, see Bowen J. Fung, Elissa Sutlief, and Marshall G. Hussain Shuler, "Dopamine and the Interdependency of Time Perception and Reward," *Neuroscience and Biobehavioral Reviews*, vol. 125 (2021): 380–391, accessed December 10, 2023, https://doi.org/10.1016/j.neubiorev.2021.02.030. Under the AIN framework, a successful prediction underlies a basic form of intrinsic valuing for any brained animal, see Karl J. Friston, Jean Daunizeau, and Stefan J. Kiebel, "Reinforcement Learning or Active Inference?," *PLOS ONE*, vol. 4, no. 7 (2009): e6421, accessed December 10, 2023, https://doi.org/10.1371/journal.pone.0006421.

[87] Thomas Nagel, "What Is It Like to Be a Bat?," *The Philosophical Review*, vol. 83, no. 4 (1974): 435–450, accessed June 1, 2023, https://doi.org/10.2307/2183914.

[88] Anil Seth, *Being You: A New Science of Consciousness* (London: Penguin Publishing Group, 2021), 11–31.

[89] Jeffrey Markowitz et al., "Spontaneous Behaviour Is Structured by Reinforcement Without Explicit Reward," *Nature*, vol. 614, no. 7946 (2023): 108–117, accessed November 20, 2023, https://doi.org/10.1038/s41586-022-05611-2.

[90] Karl Friston, "Am I Self-Conscious? (Or Does Self-Organization Entail Self-Consciousness?)," *Frontiers in Psychology*, vol. 9 (2018), accessed November 20, 2023, https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00579.

the FEP, there is room for the concept of FW – albeit a somewhat metaphorical one, given his physicalism.[91] Its origins lie in the temporal thickness (or counterfactual depth), enabling an agent to draw inferences about the consequences of her actions. Though Friston addresses this issue using two extreme examples (viruses *versus* vegans), a nonhuman mammal can be safely positioned in between. As we have argued throughout this paper, rats appear to possess the necessary cognitive ingredients to be regarded as entertaining a confined form of sophisticated agency under this description. They simulate possibilities, can monitor the accuracy of their own actions and mentally travel across time – even if minimally, up to days back and at least tens of seconds ahead – sufficiently for the demands of their *Umwelt*[92] and, presumably, to draw sophisticated counterfactual inferences. In fact, some FEP theorists (including Hohwy) explicitly refer to experiments with rodents' counterfactual reasoning when discussing advanced cognition.[93]

Moreover, a line of research on the relationship between self-control,[94] counterfactual inference, deliberation, and experience of regret leading to policy change in rodents – thus far unchallenged by deflationary explanations – demonstrates homology striking to such extent that it is utilised to explore neural and behavioural underpinnings of human deliberative processes.[95] Taken together, these findings allow us to hypothesise that rodents possess sufficient temporal thickness to experience self-evidencing based on their actions.[96] FEP theorists make this hypothesis plausible

[91] More akin to what Christopher Hitchens was talking about – "We have no other choice than to believe in FW."

[92] For an overview, see Joseph Huston and Owen Chao, "Probing the Nature of Episodic Memory in Rodents," *Neuroscience & Biobehavioral Reviews*, vol. 144 (2023): 104930, accessed November 20, 2023, https://doi.org/10.1016/j.neubiorev.2022.104930.

[93] Andrew Corcoran, Giovanni Pezzulo, and Jakob Hohwy, "From Allostatic Agents to Counterfactual Cognisers: Active Inference, Biological Regulation, and the Origins of Cognition," *Biology & Philosophy*, vol. 35, no. 3 (2020): 32, accessed November 25, 2023, https://doi.org/10.1007/s10539-020-09746-2.

[94] Many species, including rodents, entertain self-control, and this fact alone should deliver a final blow to the Cartesian view of animals as mere automatons and Frankfurt's description of them as first-order "wantons" (see for details Michael Beran, *Self-Control in Animals and People*, first edition (London: Academic Press, 2018)).

[95] Deliberation in rats: Steiner and Redish, "Behavioral and Neurophysiological Correlates of Regret in Rat Decision-Making on a Neuroeconomic Task"; Laurent and Balleine, "Factual and Counterfactual Action-Outcome Mappings Control Choice Between Goal-Directed Actions in Rats"; Hasz and Redish, "Deliberation and Procedural Automation on a Two-Step Task for Rats"; Brian M. Sweis et al., "Sensitivity to 'Sunk Costs' in Mice, Rats, and Humans," *Science*, vol. 361, no. 6398 (2018): 178–181, accessed December 17, 2023, https://doi.org/10.1126/science.aar8644. Deliberation in humans: Samantha Abram et al., "The Web-Surf Task: A Translational Model of Human Decision-Making," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 16, no. 1 (2016): 37–50, accessed November 24, 2023, https://doi.org/10.3758/s13415-015-0379-y; Abram et al., "Neural Signatures Underlying Deliberation."

[96] See Maurizio Casarrubea et al., "Detection of a Temporal Structure in the Rat Behavioural Response to an Aversive Stimulation in the Emotional Object Recognition (EOR) Task," *Physiology*

given that, to the best of our knowledge, they have not established minimal boundaries of temporal thickness for the cases of agents with the requisite minimal cognitive properties necessary for it. In fact, some FEP theorists (including Hohwy) explicitly refer to research on rodents' acumen when discussing advanced cognition.[97]

To situate rodents in the broader theoretical perspective, if we embrace a description of cognitive agents as proposed by Michael Levin,[98] we can safely assume that the cognitive light cones of rats – although significantly more modest than ours – are nevertheless more extensive than those of most cognitive agents, as they are comparable to other endothermic species.[99] Counterfactual inferencing and sophisticated self-evidencing significantly expand an agent's cognitive horizons. Suppose a rat's cognition is sufficiently powerful to uphold minimal temporal consciousness on this basis; in that case, one might term it "Embodied FW (EFW)," although, as we advised in *Of Rats and Men I*, it would be more constructive to relinquish this concept when explicating humans' and animals' core features. And if one accepts the classic human–animal dichotomy at face value, then a single species demonstrating minimal temporal consciousness with SoA would undermine the entire dichotomy, at least in this crucial aspect. For the successful advancement of such a project, there would be a missing puzzle left – the demonstration of the capacity to experience subjective perspective (SoA) as the level under which rodents perform self-evidencing, and this would be the goal for testing "intentional binding" in rats.

& Behavior, vol. 238 (2021): 113481, accessed December 18, 2023, https://doi.org/10.1016/j.physbeh.2021.113481; Eugenio Piasini et al., "Temporal Stability of Stimulus Representation Increases along Rodent Visual Cortical Hierarchies," *Nature Communications*, vol. 12, no. 1 (2021): 4448, accessed December 18, 2023, https://doi.org/10.1038/s41467-021-24456-3.

[97] See Andrew Corcoran, Giovanni Pezzulo, and Jakob Hohwy, "From Allostatic Agents to Counterfactual Cognisers: Active Inference, Biological Regulation, and the Origins of Cognition," *Biology & Philosophy*, vol. 35, no. 3 (2020): 32, accessed November 25, 2023, https://doi.org/10.1007/s10539-020-09746-2.

[98] According to Levin, the complex cognitive functions of multicellular organisms evolved through scaled-up primary drivers. These functions are facilitated by bioelectric networks already present in cells (infotaxis, bioelectric communication). Levin's theory aims to describe the cognitive abilities of diverse agents independently of their physical implementation, focusing on information processing (in the Shannon sense) and goal-directed activity. The cognitive light cone represents the spatio-temporal boundaries of events that a given system can model and attempt to influence through action. It delineates the functional boundaries (horizon) of an agent's goal space, beyond which its cognitive system is unable to operate with respect to goals. In mammals such as rats, the cognitive light cone includes notable temporal memory and local predictive abilities, indicating that their cognitive capabilities span local scales. See, Michael Levin, "The Computational Boundary of a 'Self': Developmental Bioelectricity Drives Multicellularity and Scale-Free Cognition," *Frontiers in Psychology*, vol. 10 (2019): 2688, accessed December 19, 2023, https://doi.org/10.3389/fpsyg.2019.02688.

[99] Patricia Churchland, *Conscience: The Origins of Moral Intuition*, first edition (New York: W. W. Norton & Company, 2019), 22–40; Nicolas Humphrey, *Sentience: The Invention of Consciousness* (London: The MIT Press, 2023), 148–206.

Consequently, such a finding would pave the way for gradual categories of moral agency akin to Mark Rowlands' proposition. The significance of such a demonstration, interpreted within the AIN framework, is that it would strengthen the category of the moral subject. This would be achieved owing to AIN's inclusiveness and the coherent explanation it offers of the mechanisms by which environmental causal structures are tracked. For Rowlands' conception of the moral subject, the ability to act upon embodied (affective) recognition of causal dependencies is vital, since the moral subject must be able to track and act upon them as reasons.[100] As we have tried to show by invoking research on rodents, AIN is not only inclusive and cohesive in this respect, but it also converges with the most recent empirical findings – something that cannot be said of traditional mainstream paradigms.

Conclusion

In the triptych comprising three interconnected parts, we scrutinised the concept of free will (FW) through the lens of cognitively oriented pragmatism. In the first paper, we delineated the erroneous premises underpinning the human–animal dichotomy. We linked the cultural attachment to this divide with the psychological phenomenon of anthropodenialism, reinforced by philosophical, post-Cartesian scepticism. Utilising the pragmatic maxim, we challenged widely accepted assumptions about the impact of belief in FW on human conduct, whence calling into question its indispensability. We examined various philosophical arguments related to the FW discourse – traditionally invoked to fortify the human–animal divide – and found them unconvincing. We also highlighted the limited applicability of the FW concept beyond Western societies, emphasising its cultural idiosyncrasy. In the conclusion

[100] In the concept of "moral proposition tracking," Rowlands argues that attributing a specific emotion as a reason for action involves tracking moral claims. He posits that it is justified to attribute a moral emotion to an animal's state E in a given context if and only if: 1) state E is a content-involving emotion (i.e., it is about something); 2) there exists a proposition P that expresses a moral claim; and 3) this claim must necessarily be true for the state E not be misguided. This implies that if an animal is in an emotional state appropriate to the moral context, then there exists a moral claim to which the animal responds in an adequate manner, guided by an emotion that serves as a reason for action. However, in order to establish the relationship between the attribution of tracking a moral proposition and moral action, it is first necessary to plausibly explain the animal's capacity to act based on an emotion that is not merely a non-cognitive affective reaction to an event. We deliberately tackled the problem from the cognitive angle, as within the AIN framework, emotions are understood as forms of cognition. This will be the topic will be the subject of future papers.

of the first paper, we indicated that the FW category fits the definition of anthropofabulation and postulated abandoning it as an explanatory device. Simultaneously, we pointed to the importance of the notion of agency and the potential of explaining the widespread subjective impression of action ownership (the sense of agency, SoA) underlying FW narratives. If the function of SoA is elucidated, it promises to reveal a similar kind of experience in other animals.

In the succeeding paper, we employed converging, unorthodox approaches to agency: (1) the processual metaphysics of biology, (2) the extended evolutionary synthesis narrative, (3) the free-energy principle (FEP) and its first corollary – predictive processing (PP). We aimed to reconstruct the basis of agency free from FW tenets and to facilitate the second corollary of the FEP – namely, active inference (AIN), which is further developed in the concluding paper. We argued that embracing the proposed conceptual schema – rather than the traditional rationalistic narrative – enables the uncontroversial use of folk psychology to describe animal behaviour.

In the third paper, we analysed features of advanced cognitive agents such as counterfactual inferencing and sophisticated self-evidencing, and introduced the tool for measuring SoA, called "intentional binding." Within the AIN framework, we perceive intentional binding as a form of sophisticated, embodied self-evidencing that can be captured and measured at behavioural, psychophysical, and psychophysiological levels. As such, it might be unrestricted by the human–animal dichotomy. To test this claim, we have proposed resorting to an experimental paradigm designed to measure "temporal binding" in animals. We advocated for interpreting it within the active inference conceptual framework, arguing that when supplemented with additional empirical findings on the cognitive abilities of rodents, this move might help to capture and illuminate the presence of the SoA phenomenon beyond our species. Finally, we have contextualised our proposal within Michael Levin's cognitive light cone theory, concluding with a brief discussion regarding the significance of the proposed approach for advancing theoretical discourse on non-human morality.

Throughout this three-part article, we have put forth a pragmatist perspective that is free from speciesism and transcends the metaphysically entrenched dualistic assumptions of theories invoking the concept of "the space of reasons." We have espoused a semi-compatibilist stance, as we see no need for the outright jettison of the free will concept from philosophical discourse, despite the tenuous empirical support for its ontological status. While FW belief is plausibly inconsequential at the individuals' experience level,[101] it plays a role on the societal level of institu-

[101] See Oliver Genschow et al., "Manipulating Belief in Free Will and Its Downstream Consequences: A Meta-Analysis," *Personality and Social Psychology Review: An Official Journal of the Society*

tional facts (akin to money). Concomitantly, in line with Rowlands' argumentation, we consider the problem of determinism as separate from moral responsibility – a position that could be upheld even if FW were just a story.

Be that as it may, assuming the aptness of the pragmatic approach, we must emphasise that our society's beliefs and practices, reinforced by this construct, have profound ramifications.[102] Crucially, clinging to the concept of FW in its current understanding reinforces the human–animal dichotomy, adversely impacting explanatory practices regarding other species' cognitive abilities. It also bolsters systems of exploitation of the animal kingdom, thereby exacerbating the ongoing ecological crisis. Acknowledging FW as an superficial and redundant construction and focusing on agency instead, can lead to an expansion of the human circle of moral care,[103] including broadening of the moral domain hitherto restricted to *Homo sapiens*.

Conflict of Interest Statement: The author declares that – contrary to Peter Carruthers' assertion that human sympathy is the only anchor for the moral stance of animals – he has no emotional attitude towards rodents, and has no personal acquaintance with any.

The article is a partial realisation of the PRELUDIUM 20 grant: Pragmatyczny zwrot ucieleśnionego poznania a kategoria „podmiotu motywowanego moralnie" Marka Rowlandsa (2021/41/N/HS1/04208).

This article was inspired by the cooperation with the Diverse Intelligences Summer Institute, whose programs are financed by the Templeton World Charity Foundation (grant 0333 to the University of California).

Biblioghraphy

Abram, Samantha V., Yannick-André Breton, Brandy Schmidt, A. David Redish, and Angus W. MacDonald. "The Web-Surf Task: A Translational Model of Human Decision-Making." *Cognitive, Affective, & Behavioral Neuroscience*, vol. 16, no. 1 (February 1, 2016): 37–50. https://doi.org/10.3758/s13415-015-0379-y.

for Personality and Social Psychology, Inc, vol. 27, no. 1 (2023): 52–82, accessed June 17, 2023, https://doi.org/10.1177/10888683221087527.

[102] These include the sustenance of retributive judicial systems, which result in counterproductive outcomes due to the violent norm-shaping environments into which punished individuals are cast.

[103] See Adam Waytz et al., "Ideological Differences in the Expanse of the Moral Circle," *Nature Communications*, vol. 10, no. 1 (2019): 1–12, accessed April, 14, 2024, https://doi.org/10.1038/s41467-019-12227-0.

Abram, Samantha V., Michael Hanke, A. David Redish, and Angus W. MacDonald. "Neural Signatures Underlying Deliberation in Human Foraging Decisions." *Cognitive, Affective, & Behavioral Neuroscience*, vol. 19, no. 6 (December 1, 2019): 1492–1508. https://doi.org/10.3758/s13415-019-00733-z.

Allen, Colin. "*Umwelt* or *Umwelten*? How Should Shared Representation Be Understood Given Such Diversity?" *Semiotica* 2014, no. 198 (February 1, 2014): 137–158. https://doi.org/10.1515/sem-2013-0105.

Ball, Gregory F., and Jacques Balthazart. "Evolutionary Neuroscience: Are the Brains of Birds and Mammals Really so Different?" *Current Biology*, vol. 31, no. 13 (2021): R840–42. https://doi.org/10.1016/j.cub.2021.05.004.

Ball, Philip. *The Book of Minds: Understanding Ourselves and Other Beings, from Animals to Aliens*. London: Picador, 2023.

Ball, Philip. *How Life Works: A User's Guide to the New Biology*. London: Picador, 2024.

Barrett, Lisa Feldman. "The Theory of Constructed Emotion: An Active Inference Account of Interoception and Categorization." *Social Cognitive and Affective Neuroscience*, vol. 12, no. 1 (January 1, 2017): 1–23. https://doi.org/10.1093/scan/nsw154.

Barsalou, Lawrence W. "Can Cognition Be Reduced to Action?" In *The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science*, edited by Andreas K. Engel, Karl J. Friston, and Danica Kragic, 81–96. Cambridge, Massachusetts: The MIT Press, 2016.

Barsalou, Lawrence W. "Continuity of the Conceptual System across Species." *Trends in Cognitive Sciences*, vol. 9, no. 7 (2005): 309–311. https://doi.org/10.1016/j.tics.2005.05.003.

Barsalou, Lawrence W. "Grounded Cognition." *Annual Review of Psychology*, vol. 59 (2008): 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639.

Barsalou, Lawrence W. "Simulation, Situated Conceptualization, and Prediction." *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1521 (May 12, 2009): 1281–1289. https://doi.org/10.1098/rstb.2008.0319.

Ben-Ami Bartal, Inbal. "The Complex Affective and Cognitive Capacities of Rats." *Science*, vol. 385, no. 6715 (September 20, 2024): 1298–1305. https://doi.org/10.1126/science.adq6217.

Beran, Michael. *Self-Control in Animals and People*. First edition. London: Academic Press, 2018.

Blaisdell, Aaron P., Kosuke Sawa, Kenneth J. Leising, and Michael R. Waldmann. "Causal Reasoning in Rats." *Science*, vol. 311, no. 5763 (February 17, 2006): 1020–1022. https://doi.org/10.1126/science.1121872.

Brentano, Franz. *Psychologia z Empirycznego Punktu Widzenia*. Translated by Włodzimierz Galewicz. Warszawa: PWN, 1874.

Bruineberg, Jelle. "Active Inference and the Primacy of the 'I Can'," edited by Thomas Metzinger and Wanja Wiese. *Philosophy and Predictive Processing*, vol. 5 (2017): 1–18. https://doi.org/10.155027/9783958573062.

Casarrubea, Maurizio, Manfredi Palacino, Anna Brancato, Gianluca Lavanco, Carla Cannizzaro, and Giuseppe Crescimanno. "Detection of a Temporal Structure in the Rat Behavioural Response to an Aversive Stimulation in the Emotional Object Recognition (EOR) Task." *Physiology & Behavior*, vol. 238 (September 2021): 113481. https://doi.org/10.1016/j.physbeh.2021.113481.

Chaudhry, Tuba Shahid, Sidhartha Gautam Senapati, Srikanth Gadam, Hari Priya Sri Sai Mannam, Hima Varsha Voruganti, Zainab Abbasi, Tushar Abhinav, et al. "The Impact of Microbiota on the Gut–Brain Axis: Examining the Complex Interplay and Implications." *Journal of Clinical Medicine*, vol. 12, no. 16 (2023): 5231. https://doi.org/10.3390/jcm12165231.

Churchland, Patricia S. *Conscience: The Origins of Moral Intuition*. First edition. New York: W. W. Norton & Company, 2019.

Clark, Andy. *The Experience Machine: How Our Minds Predict and Shape Reality*. New York: Pantheon, 2023.

Cole, Nicholas, Matthew Harvey, Dylan Myers-Joseph, Aditya Gilra, and Adil G. Khan. "Prediction-Error Signals in Anterior Cingulate Cortex Drive Task-Switching." *Nature Communications*, vol. 15, no. 1 (August 17, 2024): 7088. https://doi.org/10.1038/s41467-024-51368-9.

Corcoran, Andrew W, and Jakob Hohwy. "Allostasis, Interoception, and the Free Energy Principle: Feeling Our Way Forward." In *The Interoceptive Mind: From Homeostasis to Awareness*, edited by Manos Tsakiris and Helena De Preester, 272–292. Oxford: Oxford University Press, 2019.

Corcoran, Andrew W., Giovanni Pezzulo, and Jakob Hohwy. "From Allostatic Agents to Counterfactual Cognisers: Active Inference, Biological Regulation, and the Origins of Cognition." *Biology & Philosophy*, vol. 35, no. 3 (2020): 32. https://doi.org/10.1007/s10539-020-09746-2.

Courellis, Hristos S., Juri Minxha, Araceli R. Cardenas, Daniel L. Kimmel, Chrystal M. Reed, Taufik A. Valiante, C. Daniel Salzman, Adam N. Mamelak, Stefano Fusi, and Ueli Rutishauser. "Abstract Representations Emerge in Human Hippocampal Neurons during Inference." *Nature*, vol. 632 (2024): 1–9. https://doi.org/10.1038/s41586-024-07799-x.

Crawford, Elizabeth, Laura E. Knouse, Molly Kent, Dylan Vavra, Olivia Harding, Danielle LeServe, Nathan Fox, et al. "Enriched Environment Exposure Accelerates Rodent Driving Skills." *Behavioural Brain Research*, vol. 378 (January 27, 2020): 112309. https://doi.org/10.1016/j.bbr.2019.112309.

Crippen, Matthew, and Jay Schulkin. *Mind Ecologies: Body, Brain, and World*. New York: Columbia University Press, 2020.

Dennett, Daniel C. "Real Patterns." *The Journal of Philosophy*, vol. 88, no. 1 (1991): 27–51. https://ruccs.rutgers.edu/images/personal-zenon-pylyshyn/class-info/FP2012/FP2012_readings/Dennett_RealPatterns.pdf.

Dusek Jeffery A., and Howard Eichenbaum. "The Hippocampus and Memory for Orderly Stimulus Relations." *Proceedings of the National Academy of Sciences*, vol. 94, no. 13 (1997): 7109–7114. https://doi.org/10.1073/pnas.94.13.7109.

Dwyer, Dominic M., Judy Starns, and Robert C. Honey. "'Causal Reasoning' in Rats: A Reappraisal." *Journal of Experimental Psychology: Animal Behavior Processes*, vol. 35, no. 4 (2009): 578–86. https://doi.org/10.1037/a0015007.

Dwyer, Dominic M., and Michael R. Waldmann. "Beyond the Information (Not) Given: Representations of Stimulus Absence in Rats (Rattus Norvegicus)." *Journal of Comparative Psychology*, vol. 130, no. 3 (2016): 192–204. https://doi.org/10.1037/a0039733.

Fedorenko, Evelina, Anna A. Ivanova, and Tamar I. Regev. "The Language Network as a Natural Kind Within the Broader Landscape of the Human Brain." *Nature Reviews. Neuroscience*, vol. 25, no. 5 (May 2024): 289–312. https://doi.org/10.1038/s41583-024-00802-4.

Fedorenko, Evelina, Steven T. Piantadosi, and Edward A. F. Gibson. "Language Is Primarily a Tool for Communication Rather than Thought." *Nature*, vol. 630, no. 8017 (2024): 575–586. https://doi.org/10.1038/s41586-024-07522-w.

Fenton, André A. "Remapping Revisited: How the Hippocampus Represents Different Spaces." *Nature Reviews Neuroscience*, vol. 25, no. 6 (June 2024): 428–448. https://doi.org/10.1038/s41583-024-00817-x.

Fields, Chris, and Michael Levin. "How Do Living Systems Create Meaning?" *Philosophies*, vol. 5, no. 4 (2020): 1–24. https://doi.org/10.3390/philosophies5040036.

Fisher, Elizabeth L., and Jakob Hohwy. "The Universal Optimism of the Self-Evidencing Mind." *Entropy*, vol. 26, no. 6 (2024): 518. https://doi.org/10.3390/e26060518.

Friston, Karl. "Am I Self-Conscious? (Or Does Self-Organization Entail Self-Consciousness?)." *Frontiers in Psychology*, vol. 9 (2018). https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00579.

Friston, Karl. "Embodied Inference: Or 'I Think Therefore I Am, If I Am What I Think'." In *The Implications of Embodiment: Cognition and Communication*, edited by Wolfgang Tschacher and Claudia Bergomi, 89–125. Charlottesville: Imprint Academic, 2011.

Friston, Karl. "I Am Therefore I Think." In *The Unconscious: A Bridge Between Psychoanalysis and Cognitive Neuroscience*, edited by Marianne Leuzinger-Bohleber, Simon Arnold, and Mark Solms, 113–137. Psychoanalytic Explorations. London: Routledge, 2017.

Friston, Karl J., Jean Daunizeau, and Stefan J. Kiebel. "Reinforcement Learning or Active Inference?" *PLOS ONE* 4, no. 7 (July 2009): e6421. https://doi.org/10.1371/journal.pone.0006421.

Fung, Bowen J., Elissa Sutlief, and Marshall G. Hussain Shuler. "Dopamine and the Interdependency of Time Perception and Reward." *Neuroscience and Biobehavioral Reviews*, vol. 125 (2021): 380–391. https://doi.org/10.1016/j.neubiorev.2021.02.030.

Furutachi, Shohei, Alexis D. Franklin, Andreea M. Aldea, Thomas D. Mrsic-Flogel, and Sonja B. Hofer. "Cooperative Thalamocortical Circuit Mechanism for Sensory Prediction Errors." *Nature*, vol. 28 (August 2024): 1–9. https://doi.org/10.1038/s41586-024-07851-w.

Gallagher, Shaun. "Philosophical Conceptions of the Self: Implications for Cognitive Science." *Trends in Cognitive Sciences*, vol. 4, no. 1 (January 1, 2000): 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5.

Genschow, Oliver, Emiel Cracco, Jana Schneider, John Protzko, David Wisniewski, Marcel Brass, and Jonathan W. Schooler. "Manipulating Belief in Free Will and Its Downstream Consequences: A Meta-Analysis." *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, vol. 27, no. 1 (February 2023): 52–82. https://doi.org/10.1177/10888683221087527.

Griffin, Donald R. "From Cognition to Consciousness." *Animal Cognition*, vol. 1, no. 1 (July 1, 1998): 3–16. https://doi.org/10.1007/s100710050002.

Haggard, Patrick. "Sense of Agency in the Human Brain." *Nature Reviews Neuroscience*, vol. 18, no. 4 (April 2017): 196–207. https://doi.org/10.1038/nrn.2017.14.

Haggard, Patrick, Sam Clark, and Jeri Kalogeras. "Voluntary Action and Conscious Awareness." *Nature Neuroscience*, vol. 5, no. 4 (April 2002): 382–385. https://doi.org/10.1038/nn827.

Hasz, Brendan M., and A. David Redish. "Deliberation and Procedural Automation on a Two-Step Task for Rats." *Frontiers in Integrative Neuroscience*, vol. 12 (2018): 30. https://doi.org/10.3389/fnint.2018.00030.

Hateren, Hans van. "Constructing a Naturalistic Theory of Intentionality." *Philosophia*, vol. 49 (2021). https://doi.org/10.1007/s11406-020-00255-w.

Hawkins, Jeff. *A Thousand Brains: A New Theory of Intelligence*. New York: Hachette Book Group, 2021.

Hawkins, Jeff, and Subutai Ahmad. "Why Neurons Have Thousands of Synapses, a Theory of Sequence Memory in Neocortex." *Frontiers in Neural Circuits*, vol. 10 (March 30, 2016). https://doi.org/10.3389/fncir.2016.00023.

Hawkins, Jeff, Marcus Lewis, Mirko Klukas, Scott Purdy, and Subutai Ahmad. "A Framework for Intelligence and Cortical Function Based on Grid Cells in the Neocortex." *Frontiers in Neural Circuits*, vol. 12 (2018): 121. https://doi.org/10.3389/fncir.2018.00121.

Hinwar, Rish P., and Anthony J. Lambert. "Anauralia: The Silent Mind and Its Association With Aphantasia." *Frontiers in Psychology*, vol. 12 (2021). https://doi.org/10.3389/fpsyg.2021.744213.

Hohwy, Jakob. "Conscious Self-Evidencing." *Review of Philosophy and Psychology*, vol. 13, no. 4 (December 1, 2022): 809–828. https://doi.org/10.1007/s13164-021-00578-x.

Humphrey, Nicolas. *Sentience: The Invention of Consciousness*. London: The MIT Press, 2023.

Hurlburt, Russell T., and Sarah A. Akhter. "Unsymbolized Thinking." *Consciousness and Cognition*, vol. 17, no. 4 (December 1, 2008): 1364–1374. https://doi.org/10.1016/j.concog.2008.03.021.

Huston, Joseph P., and Owen Y. Chao. "Probing the Nature of Episodic Memory in Rodents." *Neuroscience & Biobehavioral Reviews*, vol. 144 (January 1, 2023): 104930. https://doi.org/10.1016/j.neubiorev.2022.104930.

Johnson, Mark L., and Jay Schulkin. *Mind in Nature: John Dewey, Cognitive Science, and a Naturalistic Philosophy for Living*. Cambridge, Massachusetts: The MIT Press, 2023.

Kirchhoff, Michael, Thomas Parr, Ensor Palacios, Karl Friston, and Julian Kiverstein. "The Markov Blankets of Life: Autonomy, Active Inference and the Free Energy Principle." *Journal of The Royal Society Interface*, vol. 15, no. 138 (January 2018): 20170792. https://doi.org/10.1098/rsif.2017.0792.

Kononowicz, Tadeusz Władysław, Virginie van Wassenhove, and Valérie Doyère. "Rodents Monitor Their Error in Self-Generated Duration on a Single Trial Basis." *Proceedings of the National Academy of Sciences*, vol. 119, no. 9 (March 2022): e2108850119. https://doi.org/10.1073/pnas.2108850119.

Kull, Kalevi, Dmitry Kozhevnikov, and Magdalena Kozhevnikova. "O logice zwierzęcych umweltów. Subiektywna teraźniejszość zwierząt, czyli zoosemiotyka wyboru i uczenia się." *Zoophilologica. Polish Journal of Animal Studies*, vol. 11, no. 1 (2023): 1–20. https://doi.org/10.31261/ZOOPHILOLOGICA.2023.11.06.

Lai, Chongxi, Shinsuke Tanaka, Timothy D. Harris, and Albert K. Lee. "Volitional Activation of Remote Place Representations with a Hippocampal Brain–Machine Interface." *Science*, vol. 382, no. 6670 (2023): 566–573. https://doi.org/10.1126/science.adh5206.

Laurent, Vincent, and Bernard W. Balleine. "Factual and Counterfactual Action-Outcome Mappings Control Choice between Goal-Directed Actions in Rats." *Current Biology*, vol. 25, no. 8 (2015): 1074–1079. https://doi.org/10.1016/j.cub.2015.02.044.

Lee, David G., Caroline A. McLachlan, Ramon Nogueira, Osung Kwon, Alanna E. Carey, Garrett House, Gavin D. Lagani, Danielle LaMay, Stefano Fusi, and Jerry L. Chen. "Perirhinal Cortex Learns a Predictive Map of the Task Environment." *Nature Communications*, vol. 15, no. 1 (July 2, 2024): 5544. https://doi.org/10.1038/s41467-024-47365-7.

Leising, Kenneth J., Jared Wong, Michael R. Waldmann, and Aaron P. Blaisdell. "The Special Status of Actions in Causal Reasoning in Rats." *Journal of Experimental Psychology: General*, vol. 137, no. 3 (2008): 514–527. https://doi.org/10.1037/0096-3445.137.3.514.

Leptourgos, Pantelis, and Philip R. Corlett. "Embodied Predictions, Agency, and Psychosis." *Frontiers in Big Data*, vol. 3 (2020). https://www.frontiersin.org/articles/10.3389/fdata.2020.00027.

Levin, Michael. "Bioelectric Networks: The Cognitive Glue Enabling Evolutionary Scaling from Physiology to Mind." *Animal Cognition*, May 19, 2023. https://doi.org/10.1007/s10071-023-01780-3.

Levin, Michael. "The Computational Boundary of a 'Self': Developmental Bioelectricity Drives Multicellularity and Scale-Free Cognition." *Frontiers in Psychology*, vol. 10 (December 13, 2019): 2688. https://doi.org/10.3389/fpsyg.2019.02688.

Limanowski, Jakub, and Felix Blankenburg. "Minimal Self-Models and the Free Energy Principle." *Frontiers in Human Neuroscience*, vol. 7 (2013). https://www.frontiersin.org/articles/10.3389/fnhum.2013.00547.

Linson, Adam, Andy Clark, Subramanian Ramamoorthy, and Karl Friston. "The Active Inference Approach to Ecological Perception: General Information Dynamics for Natural and Artificial Embodied Cognition." *Frontiers in Robotics and AI*, vol. 5 (2018). https://www.frontiersin.org/articles/10.3389/frobt.2018.00021.

Luu, Phan, Don M. Tucker, and Karl Friston. "From Active Affordance to Active Inference: Vertical Integration of Cognition in the Cerebral Cortex Through Dual Subcortical Control Systems." *Cerebral Cortex*, vol. 34, no. 1 (January 1, 2024): bhad458. https://doi.org/10.1093/cercor/bhad458.

Magnani, Lorenzo. "Animal Abduction." In *Abductive Cognition: The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning*, edited by Lorenzo Magnani, 265–316. Cognitive Systems Monographs. Berlin, Heidelberg: Springer, 2009. https://doi.org/10.1007/978-3-642-03631-6_5.

Markowitz, Jeffrey E., Winthrop F. Gillis, Maya Jay, Jeffrey Wood, Ryley W. Harris, Robert Cieszkowski, Rebecca Scott, et al. "Spontaneous Behaviour Is Structured by Reinforcement without Explicit Reward." *Nature*, vol. 614, no. 7946 (February 2023): 108–17. https://doi.org/10.1038/s41586-022-05611-2.

Masís, Javier, Travis Chapman, Juliana Y Rhee, David D. Cox, and Andrew M Saxe. "Strategically Managing Learning during Perceptual Decision Making," edited by Valentin Wyart, Michael J Frank, and Konstantinos Tsetsos. *eLife*, vol. 12 (2023): e64978. https://doi.org/10.7554/eLife.64978.

McMillen, Patrick, and Michael Levin. "Collective Intelligence: A Unifying Concept for Integrating Biology Across Scales and Substrates." *Communications Biology*, vol. 7, no. 1 (2024): 378. https://doi.org/10.1038/s42003-024-06037-4.

Miller, Mark, and Andy Clark. "Happily Entangled: Prediction, Emotion, and the Embodied Mind." *Synthese*, vol. 195, no. 6 (June 1, 2018): 2559–2575. https://doi.org/10.1007/s11229-017-1399-7.

Misak, Cheryl. "The Metaphysical Club." In *The Routledge Companion to Pragmatism*, edited by Scott F. Aikin and Robert B. Talisse, 7–12. New York–London: Routledge, 2023.

Mitchell, Kevin J. *Free Agents: How Evolution Gave Us Free Will*. First edition. Princeton Oxford: Princeton University Press, 2023.

Monzel, Merlin, David Mitchell, Fiona Macpherson, Joel Pearson, and Adam Zeman. "Aphantasia, Dysikonesia, Anauralia: Call for a Single Term for the Lack of Mental Imagery–Commentary on Dance et al. (2021) and Hinwar and Lambert (2021)." *Cortex*, vol. 150 (2022): 149–152. https://doi.org/10.1016/j.cortex.2022.02.002.

Mountcastle, Vernon B. "An Organizing Principle for Cerebral Function: The Unit Module and the Distributed System." In *The Mindful Brain: Cortical Organization and the Group-Selective Theory of Higher Brain Function*, edited by Gerald M. Edelman and Vernon B. Moundcastle, 7–50. Cambridge: The MIT Press, 1978.

Nagel, Thomas. "What Is It Like to Be a Bat?" *The Philosophical Review*, vol. 83, no. 4 (1974): 435–450. https://doi.org/10.2307/2183914.

Nave, Kathryn. *A Drive to Survive: The Free Energy Principle and the Meaning of Life*. London: The MIT Press, 2025.

Nave, Kathryn, George Deane, Mark Miller, and Andy Clark. "Wilding the Predictive Brain." *WIREs Cognitive Science*, vol. 11, no. 6 (2020): e1542. https://doi.org/10.1002/wcs.1542.

Nedergaard, Johanne S. K., and Gary Lupyan. "Not Everybody Has an Inner Voice: Behavioral Consequences of Anendophasia." *Psychological Science*, vol. 35(7) (2024): 09567976241243004. https://doi.org/10.1177/09567976241243004.

Nicholson, Daniel J. "Reconceptualizing the Organism From Complex Machine to Flowing Stream." In *Everything Flows: Towards a Processual Philosophy of Biology*, edited by Daniel J. Nicholson and John Dupré, First edition, 139–166. Oxford: Oxford University Press, 2018.

Nieh, Edward H., Manuel Schottdorf, Nicolas W. Freeman, Ryan J. Low, Sam Lewallen, Sue Ann Koay, Lucas Pinto, Jeffrey L. Gauthier, Carlos D. Brody, and David W. Tank. "Geometry of Abstract Learned Knowledge in the Hippocampus." *Nature*, vol. 595, no. 7865 (July 2021): 80–84. https://doi.org/10.1038/s41586-021-03652-7.

Parr, Thomas, Giovanni Pezzulo, and Karl J. Friston. *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. New York: The MIT Press, 2022.

Peña-Guzmán, David M. *When Animals Dream: The Hidden World of Animal Consciousness*. Princeton–Oxford: Princeton University Press, 2022.

Pezzulo, Giovanni, Emilio Cartoni, Francesco Rigoli, Léo Pio-Lopez, and Karl Friston. "Active Inference, Epistemic Value, and Vicarious Trial and Error." *Learning & Memory*, vol. 23, no. 7 (2016): 322–338. https://doi.org/10.1101/lm.041780.116.

Piasini, Eugenio, Liviu Soltuzu, Paolo Muratore, Riccardo Caramellino, Kasper Vinken, Hans Op de Beeck, Vijay Balasubramanian, and Davide Zoccolan. "Temporal Stability of Stimulus Representation Increases along Rodent Visual Cortical Hierarchies." *Nature Communications*, vol. 12, no. 1 (2021): 4448. https://doi.org/10.1038/s41467-021-24456-3.

Potter, Henry D., and Kevin J. Mitchell. "Naturalising Agent Causation." *Entropy (Basel, Switzerland)*, vol. 24, no. 4 (2022): 1–18. https://doi.org/10.3390/e24040472.

Ramstead, Maxwell. "The Free Energy Principle – a Precis." *Dialectical Systems* (2023). https://doi.org/10.5281/zenodo.10014870.

Ramstead, Maxwell J. D., Karl J. Friston, and Inês Hipólito. "Is the Free-Energy Principle a Formal Theory of Semantics? From Variational Density Dynamics to Neural and Phenotypic Representations." *Entropy*, vol. 22, no. 8 (2020): 1–29. https://doi.org/10.3390/e22080889.

Redish, A. David. "Vicarious Trial and Error." *Nature Reviews Neuroscience*, vol. 17, no. 3 (2016): 147–159. https://doi.org/10.1038/nrn.2015.30.

Rouleau, Nicolas, and Michael Levin. "Brains and Where Else? Mapping Theories of Consciousness to Unconventional Embodiments." *OSF Preprints*, 2025, 1–32. https://doi.org/10.31234/osf.io/va5mk.

Rowlands, Mark. *Can Animals Be Moral?* Reprint edition. Oxford University Press, 2012.

Sawa, Kosuke. "Predictive Behavior and Causal Learning in Animals and Humans." *Japanese Psychological Research*, vol. 51, no. 3 (2009): 222–233. https://doi.org/10.1111/j.1468-5884.2009.00396.x.

Schultz, Wolfram. "Dopamine Reward Prediction Error Coding." *Dialogues in Clinical Neuroscience*, vol. 18, no. 1 (2016): 23–32. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4826767/.

Schultz, Wolfram, Peter Dayan, and Pendleton Read Montague. "A Neural Substrate of Prediction and Reward." *Science*, vol. 275, no. 5306 (March 14, 1997): 1593–1599. https://doi.org/10.1126/science.275.5306.1593.

Seth, Anil. *Being You: A New Science of Consciousness*. London: Penguin Publishing Group, 2021.

Seth, Anil K. "Being a Beast Machine: The Origins of Selfhood in Control-Oriented Interoceptive Inference." In *Andy Clark and His Critics*, edited by Matteo Colombo, Elizabeth Irvine, and Mog Stapleton, 238–253. New York: Oxford University Press, 2019.

Seth, Anil K. "Inference to the Best Prediction." In *Open MIND*, edited by Thomas Metzinger and Jennifer M. Windt. Frankfurt am Main: MIND Group, 2015.

Seth, Anil K. "Interoceptive Inference, Emotion, and the Embodied Self." *Trends in Cognitive Sciences*, vol. 17, no. 11 (2013): 565–573. https://doi.org/10.1016/j.tics.2013.09.007.

Seth, Anil K., and Karl J. Friston. "Active Interoceptive Inference and the Emotional Brain." *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 371, no. 1708 (2016): 20160007. https://doi.org/10.1098/rstb.2016.0007.

Silva, Francisco J., and Kathleen M. Silva. "Causal Reasoning in Rats (Blaisdell et al. 2006)." In *Encyclopedia of Evolutionary Psychological Science*, edited by Todd K. Shackelford and Viviana A. Weekes-Shackelford, 936–940. Cham: Springer International Publishing, 2021. https://doi.org/10.1007/978-3-319-19650-3_3116.

Słuszkiewicz, Zbigniew. "Of Rats and Men I: A Pragmatist Take on the Concept of Free Will as a Challenge to the Human-Animal Dichotomy." *Zoophilologica. Polish Journal of Animal Studies*, vol. 2, no. 14 (2024): 1–43. https://doi.org/10.31261/ZOOPHILOLOGICA.2024.14.12.

Słuszkiewicz, Zbigniew. "Szczur w badaniach psychologii społecznej i w naukach kognitywnych. Bestia sacer czy istota motywowana moralnie?" *Zoophilologica. Polish Journal of Animal Studies*, vol. 10, no. 2 (2022): 1–70. https://doi.org/10.31261/ZOOPHILOLOGICA.2022.10.04.

Smith, Ryan, Karl J. Friston, and Christopher J. Whyte. "A Step-by-Step Tutorial on Active Inference and Its Application to Empirical Data." *Journal of Mathematical Psychology*, vol. 107 (2022): 102632. https://doi.org/10.1016/j.jmp.2021.102632.

So, Wing Yi, Karl J. Friston, and Victorita Neacsu. "The Inherent Normativity of Concepts." *Minds and Machines*, vol. 34, no. 4 (2024): 1–21. https://doi.org/10.1007/s11023-024-09697-7.

Solms, Mark. "The Hard Problem of Consciousness and the Free Energy Principle." *Frontiers in Psychology*, vol. 9 (2019). https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02714.

Solms, Mark. *The Hidden Spring: A Journey to the Source of Consciousness*. New York: W. W. Norton & Company, 2021.

Solms, Mark, and Karl Friston. "How and Why Consciousness Arises: Some Considerations from Physics and Physiology." *Journal of Consciousness Studies*, vol. 25 (2018): 202–238. https://discovery.ucl.ac.uk/id/eprint/10057681/1/Friston_Paper.pdf?ref=quillette.com.

Sprevak, Mark, and Ryan Smith. "An Introduction to Predictive Processing Models of Perception and Decision-Making." *Topics in Cognitive Science*, vol. 00 (2023): 1–28. https://doi.org/10.1111/tops.12704.

Steiner, Adam P., and A. David Redish. "Behavioral and Neurophysiological Correlates of Regret in Rat Decision-Making on a Neuroeconomic Task." *Nature Neuroscience*, vol. 17, no. 7 (July 2014): 995–1002. https://doi.org/10.1038/nn.3740.

Sweis, Brian M., Samantha V. Abram, Brandy J. Schmidt, Kelsey D. Seeland, Angus W. MacDonald, Mark J. Thomas, and A. David Redish. "Sensitivity to 'Sunk Costs' in Mice, Rats, and Humans." *Science*, vol. 361, no. 6398 (July 13, 2018): 178–181. https://doi.org/10.1126/science.aar8644.

Tramacere, Antonella, and Colin Allen. "Temporal Binding: Digging into Animal Minds through Time Perception." *Synthese*, vol. 200, no. 1 (2022): 1–24. https://doi.org/10.1007/s11229-022-03456-w.

Tschantz, Alexander, Laura Barca, Domenico Maisto, Christopher Buckley, Anil Seth, and Giovanni Pezzulo. "Simulating Homeostatic, Allostatic and Goal-Directed Forms of Interoceptive Control Using Active Inference." *Biological Psychology*, vol. 169 (2022). https://doi.org/10.1016/j.biopsycho.2022.108266.

Vasilev, Dmitrii, Isabel Raposo, and Nelson K Totah. "Brightness Illusions Evoke Pupil Constriction Preceded by a Primary Visual Cortex Response in Rats." *Cerebral Cortex*, vol. 33, no. 12 (June 15, 2023): 7952–7959. https://doi.org/10.1093/cercor/bhad090.

Vitti Rodrigues, Mariana, and Claus Emmeche. "Abduction: Can Non-Human Animals Make Discoveries?" *Biosemiotics*, vol. 10 (July 1, 2017): 295–313. https://doi.org/10.1007/s12304-017-9300-0.

Waldmann, Michael R., Martina Schmid, Jared Wong, and Aaron P. Blaisdell. "Rats Distinguish between Absence of Events and Lack of Evidence in Contingency Learning." *Animal Cognition*, vol. 15, no. 5 (2012): 979–990. https://doi.org/10.1007/s10071-012-0524-8.

Wang, Qinwen, Qianyue Yang, and Xingyin Liu. "The Microbiota–Gut–Brain Axis and Neurodevelopmental Disorders." *Protein & Cell*, vol. 14, no. 10 (2023): 762–775. https://doi.org/10.1093/procel/pwad026.

Waytz, Adam, Ravi Iyer, Liane Young, Jonathan Haidt, and Jesse Graham. "Ideological Differences in the Expanse of the Moral Circle." *Nature Communications*, vol. 10, no. 1 (September 26, 2019): 1–12. https://doi.org/10.1038/s41467-019-12227-0.

Yokose, Jun, William D. Marks, and Takashi Kitamura. "Visuotactile Integration Facilitates Mirror-Induced Self-Directed Behavior through Activation of Hippocampal Neuronal Ensembles in Mice." *Neuron*, vol. 112, no. 2 (January 17, 2024): 306–318.e8. https://doi.org/10.1016/j.neuron.2023.10.022.

Zeithamova, Dagmar, Margaret L. Schlichting, and Alison R. Preston. "The Hippocampus and Inferential Reasoning: Building Memories to Navigate Future Decisions." *Frontiers in Human Neuroscience*, vol. 6 (2012). https://doi.org/10.3389/fnhum.2012.00070.

Zimmerman, Aaron. "Pragmatism in the Philosophy of Mind." In *The Routledge Companion to Pragmatism*, edited by Scott F. Aikin and Robert B. Talisse, first edition, 226–238. New York–London: Routledge, 2023.

Zbigniew Słuszkiewicz is a graduate in Philosophy from the Pedagogical University of Cracow (2007). He also completed postgraduate studies in Special Pedagogy at the same institution (2012); Autism, Asperger Syndrome, and Overall Developmental Disorders: Diagnosis, and Therapy (with distinction) at the WSB University of Dąbrowa Górnicza (2015); and Psychology of Crisis and Crisis Intervention at the Jagiellonian University of Cracow (2019). He is currently a student of Doctoral School of UKEN, Cracow, and an alumnus of the Diverse Intelligences Summer Institute 2021, which was supported by the Templeton Foundation grant and conducted by the University of California, Los Angeles. In 2022, he was awarded a PRELUDIUM scientific grant from Narodowe Centrum Nauki for a research program titled "The Pragmatic Turn in Embodied Cognition and Mark Rowlands' Category of 'the Moral Subject'." He is an ethics teacher and behavioural therapist at the Centrum Autyzmu i Całościowych Zaburzeń Rozwojowych in Cracow. E-mail: zbigniew.sluszkiewicz@doktorant.up.krakow.pl.

Zbigniew Słuszkiewicz jest absolwentem filozofii na Uniwersytecie Pedagogicznym w Krakowie (2007). Ukończył także studia podyplomowe z zakresu Pedagogiki Specjalnej na tej samej uczelni (2012), Autyzmu, Zespołu Aspergera i Ogólnych Zaburzeń Rozwojowych, Diagnostyki i Terapii (z wyróżnieniem) w Wyższej Szkole Bankowej w Dąbrowie Górniczej (2015) oraz Psychologii Kryzysu i Interwencji Kryzysowej na Uniwersytecie Jagiellońskim (2019). Obecnie doktorant UKEN w Krakowie. Jest absolwentem Diverse Intelligences Summer Institute (2021), wspieranego przez grant Fundacji Johna Templetona i prowadzonego przez Uniwersytet Kalifornijski w Los Angeles. W 2022 roku otrzymał grant naukowy PRELUDIUM Narodowego Centrum Nauki na dwuletni program badawczy pt. „Zwrot pragmatyczny w poznaniu ucieleśnionym oraz kategoria »podmiotu moralnego« Marka Rowlandsa". Zbigniew pracuje jako nauczyciel etyki i terapeuta behawioralny w Centrum Autyzmu i Całościowych Zaburzeń Rozwojowych w Krakowie. E-mail: zbigniew.sluszkiewicz@doktorant.up.krakow.pl.