doi 10.18699/bgrs2024-10-19

Toward human-like intelligence in artificial systems: the feeling of knowing

Kralik J.D.*

Department of Brain and Cognitive Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea

** jerald.kralik@gmail.com*

Key words: artificial general intelligence; affect; subjective experience; neuroscience; cell signalling

Motivation and Aim: Human intelligence includes a sense of knowing. Patterns and probabilities are important (and used by the brain), but they result from something more to the point: the reasons why they are observed, what they result from. And human-like knowing cares about these underlying reasons. That is, it is making sense of things (see [1]). And to make sense, this theoretical analysis argues, requires not only determining such things as *what* is happening (e. g., what or who is involved), *why* (the causal agents), *how* (mechanism of change), the *expected outcome*, and *evaluating* it (typically with respect to valence of good or bad), but also the *feeling* one has when they believe they have understood the situation properly. The *first aim* of the current report is to clarify this 'feeling of knowing' and to show how an evolutionary analysis reveals its significance for human intelligence. The *second aim* is then to circumscribe its possible origins in humans, and thereby provide a framework for more detailed computational and biological work to (a) determine exactly how it emerges and influences human intelligence, and in turn, (b) model it for integration in artificial systems.

Methods: The theoretical analysis takes as premise that the *feeling of knowing* is of the same essential character as general affect, emotion, care, and even subjective awareness underpinning consciousness. Based on this, I attempt to examine and outline all possible sources of the feeling of knowing, highlighting the main findings in affective neuroscience (such as structures in the limbic system), but then zooming out with an evolutionary and biological perspective: considering cells of single- (both prokaryote, like bacteria, and eukaryote, like algae) and multicellular organisms; systems of eukaryote cells in multicellular organisms such as ourselves (pointing to intra and extra cellular coordination and signaling); the uniqueness of neural signals of the nervous system; and the scale and organization (including hierarchy and loops) achieved via layers of signals reflecting *need/affect/care* (for survival and reproduction and beyond), that give rise to the sense of care and feeling in other animals, and ultimately in true human understanding.

Results: Four key considerations each led to critical insights. First, to address the question of how significant the 'feeling of knowing' is for higher intelligence, we look to the relevant research in psychology and neuroscience. The literature thus far, however, provides mixed results on the extent *cognition,* that is, knowledge and thinking, can be separated from affect (and feeling). An evolutionary approach can provide insight. And to do this requires considering the origins of the most foundational *sensing-cognition-action cycle* in the history of life: that is, with individual (prokaryotic and then eukaryotic) cells, and then the progression to multicellular organisms. Put differently, we can examine which came first – sensing, action, cognition, or need – and then how

these general components interact. **Insight one:** The most foundational factor is *need* (for resources and homeostasis), followed by *need-driven action* (to attain what is needed), which then requires *sensory guidance*, with continued elaboration of sensory, action, and need processing via *cognition*. And importantly, all elements of this fundamental cycle (need, action, sensation, cognition) appear tightly coupled (i. e., recurrent rather than simply sequential), with for example, action anticipation influencing sensory processing.

The next question is whether the need-sensation-cognition-action cycle components remain tightly coupled as one ascends from simpler (e. g., cells to nervous systems, brainstem to hypothalamus) to more complex regions (sensorimotor cortex to the highest association areas), especially in humans. **Insight two:** tight coupling should and does appear to occur at all levels of the human brain. Focusing on need or affect specifically, although some affective neuroscientists have provided evidence for affective influences at the highest levels of the human brain and thinking, the extent of the influence remains underappreciated [1, 2]. But as suggested by evolutionary psychology and biology, and by the specialized nuclei in lower and mid brain structures (e. g., hypothalamus), the human brain appears to be organized around the fundamental *needs* of life (e. g., resource attainment, especially fluid and food, defense, mating, etc.), with affective-emotional instinct and knowledge priors producing specializations even at the highest levels of cortex: with the clearest higher-level example being the so-called social brain [1].

Is the 'feeling of knowing', then, essentially synonymous with affect and reward, providing the drive (wanting) and payoff (liking) for goal attainment? **Insight three**: It is, of course, more than this. Nonetheless, it is clear that for all aspects of intelligence – such as knowledge seeking, attaining and utilizing (e. g., in reasoning) – affective-emotional-subjective influences, elaborated into higher-level versions, are involved: such as the sense of delight, empowerment, prestige, and satisfaction with problems solved, knowledge gaps filled, discoveries made [3]. And a strong implication of the current analysis suggests that without such sensations, such as the relief and satisfaction one feels when a correct hypothesis turns confusion to clarity and broader understanding, true knowing and belief remain incomplete.

Finally, where and how does this 'feeling of knowing' originate in the brain – how does it emerge? The answer indeed remains out-of-reach, and so requires (a) a description of what is currently known; and (b) a roadmap leading to a comprehensive understanding. **Insight four:** there are many possible sources producing emergent subjective experience, which together construct such a framework and roadmap. For humans, and vertebrates more generally, *limbic* brain structures have been especially implicated [2]. More specifically, there is strong, compelling evidence for the conversion of neural signals representing sensory information (e. g., higher firing rate with larger food quantity: large chocolate bar!) to subjective experience (e. g., lower firing rate with high food quantity due to satiation: no more chocolate!) in these limbic structures. Prominent examples include the orbitofrontal cortex (taste), the nucleus accumbens of the subcortical basal ganglia (wanting and liking), and the insula cortex (pain) [1]. But zooming in remains unclear. For example, for *reward*, *wanting* and *liking* components have been tied to dopamine and endogenous opioid infusion (from neurons in the midbrain) into the nucleus accumbens, respectively [4, 5]. And so neurochemistry is critical. But what is happening? The neurochemicals open and close channels to excite and inhibit various neural populations. But how subjective experience derives or emerges from the influence of such infusions is unknown. Zooming out suggests a potentially wider influence of ‘cortical, basal ganglia, thalamus, back to cortex’ loop structures (let alone the bottom-up influence of the hypothalamus and midbrain as well) [1, 2, 6, 7]. And so the subjective experience of the feeling of knowing potentially results from the recurrence of these loops and/or from other aspects of the greater complexity of this larger system. Critically, however, a great deal of evidence suggests that other organisms at much simpler levels of nervous system complexity also show signs of affect and subjective experience [8]. And so comparative work should also help. Indeed, returning to the prior evolutionary examination, there is *need* at all levels: i. e., inside every cell, with organelles and other cellular structures sending *signals* to communicate this need. For any given cell, let alone the multicellularity of even very simple organisms, tens of thousands to millions of signals lead to activation of homeostasis systems and affect. Whether the critical factor for emergence results from the specific character of the neural signal (together with activating neurochemistry) without significant influence from the massive *scale* (and hierarchical organization) of nonneural cellular signals remains an open question. Examination on all fronts is warranted. This importantly includes more computational work, to help clarify all roles and continue circumscribing and triangulating the source(s).

Conclusion: One of the most vexing and intractable elements of human intelligence is in fact an affective-type component: not only the desire to know, but the sensation – or feeling – of knowing; which is believed and argued here to be required for true human intelligence, and therefore for comparable artificial systems. My theoretical analysis, together with other theoretical and empirical work, circumscribes a great deal of possible sources leading to the emergence of this subjective experience, ranging from the sheer scale and extraordinary computational complexity in layered and recurrent architectures, to potential specific aspects of the actual chemistry – though the latter possibility remains underspecified, as it is unclear what of the chemistry could not be artificially reproduced. Although the hardware in biological and most artificial systems (like computers) are different, they nevertheless share fundamental-force underpinnings (electromagnetic, strong, weak). Nevertheless, the insights obtained here, though daunting, still suggest significant progress can be made, with concerted efforts along multiple lines of inquiry, toward understanding and thereby ultimately reproducing ‘feeling’ for true human-like knowing in artificial systems.

References

1. Gazzaniga M.S., Ivry R.B., Mangun G.R. Cognitive neuroscience: the biology of the mind. WW Norton & Company, 2018
2. LeDoux J.E. Emotion Circuits in the Brain. *Ann Rev Neurosci*. 2003;23(1):155-184
3. Loewenstein G. The psychology of curiosity: A review and reinterpretation. *Psychol Bull.* 1994;116(1):75-98. doi 10.1037/0033-2909.116.1.75
4. Morales I., Berridge K.C. ‘Liking’ and ‘wanting’ in eating and food reward: Brain mechanisms and clinical implications. *Physiol Behav*. 2020;227:113152. doi 10.1016/j.physbeh.2020.113152
5. Jang H., Jung K., Jeong J., Park S.K., Kralik J.D., Jeong J. Nucleus accumbens shell moderates preference bias during voluntary choice behavior. *Social Cognit Affective Neurosci*. 2017;12(9):1428-1436. doi 10.1093/scan/nsx072
6. Kralik J.D. Architectural Design of Mind & Brain from an Evolutionary Perspective. AAAI Fall Symposium: Standard Model of Mind, 2017
7. Kralik J.D. Toward a Comprehensive List of Necessary Abilities for Human Intelligence, Part 1: Constructing Knowledge. In: Proceedings of the 15th International Conference on Artificial General Intelligence (AGI 2022). 2022;271-281. doi 10.1007/978-3-031-19907-3_25
8. Koch C. The Quest for Consciousness: A Neurobiological Approach. Roberts & Co, 2004