The Schema Spectrum: Explicit, Implicit, and Emergent Structures in AI and the Brain

Mandana Samiei[1,2], Doina Precup[1,2,3,7], and Blake A. Richards[1,2,3,4,5,6]

[1]Mila - Quebec Artificial Intelligence Institute, Montreal, QC, Canada
[2]School of Computer Science, McGill University, Montreal, QC, Canada
[3]CIFAR Learning in Machines and Brains Program, Toronto, ON, Canada
[4]Department of Neurology & Neurosurgery, McGill University, Montreal, QC, Canada
[5]Montreal Neurological Institute, McGill University, Montreal, QC, Canada
[6]Google, Paradigms of Intelligence Team
[7]Google DeepMind
[*]Correspondence: mandana.samiei@mail.mcgill.ca

SUMMARY

There is a long history of interplay between the brain sciences and AI in the area of schema theory. Schemas are typically defined as abstract mental structures that capture how events unfold across contexts. Evidence suggests that the brain relies on schemas to interpret and encode new information. Classical models have treated schemas as high-level abstract structures, distinct from detail-rich episodic memories. There are two assumptions buried within this classical approach: (1) that schemas are explicitly represented in the brain, and (2) that schemas are categorically different from episodic memories. Motivated by recent advances in generative AI, we challenge these two assumptions. First, we propose that schemas may not exist as explicit entities in the brain, but rather, they serve as a conceptual framework for describing how existing knowledge stored in distributed networks can impact downstream information processing and learning. Second, we suggest that schematization exists along a continuum, with memories and knowledge varying gradually in their level of abstraction and specificity. This perspective motivates new experimental approaches for probing memory schematization, and offers fresh avenues for leveraging schema theory in AI.

KEYWORDS

Schema Theory, Episodic Memory, Semantic Memory, Memory Schematization, Replay, Memory Consolidation, Distributed Representations, Knowledge Abstraction, Predictive Processing, Mental Models, Neuro-AI

Introduction to Schema Theory

Schema theory has a long and influential history across psychology, cognitive science, and neuroscience[1–6]. It also played a pivotal role in early artificial intelligence (AI) research[7,8]. The concept of schemas was originally introduced by Piaget[1] in his work on cognitive development. For Piaget, schemas were frameworks for knowledge—cognitive structures that allow individuals to organize and interpret information as they develop. He proposed that learning occurs through two complementary processes[1,9]. The process of **assimilation** refers to situations where new experiences or knowledge are re-interpreted to fit existing schemas. For example, a child with a schema for "dogs" may refer to any furry four-legged creature as a dog. In contrast, **accommodation** refers to situations where existing schemas need to be modified to account for novel information. For example, a child learning that furry four-legged creatures that meow are not dogs, but cats, will modify their schema of furry four-legged animals. The dynamic adaptation of schemas remains a cornerstone of developmental psychology[10].

Frederic Bartlett[2] extended the concept of schemas into the domain of memory research. In his classic studies on recall, Bartlett demonstrated that individuals reconstruct memories based on pre-existing knowledge frameworks rather than retrieving exact replicas of past experiences. His experiments on story recall showed that participants unconsciously reshaped narratives to fit their cultural and cognitive expectations, often omitting or altering details that did not align with their schemas. For example, people of a European background who read an indigenous North American text would recall that the characters went "fishing" or "sailing" even though text stated that they went "seal hunting". Bartlett concluded from these studies that memory recall is, in part, a reconstructive process that depends on our schemas, and not an exact recall of past information. This idea foreshadowed later discussions on memory distortion, constructive retrieval, and the role of schemas in shaping perception and recall[11].

By the late 20th century, schema theory was formalized in cognitive science through models of structured knowledge representation. Roger Schank and Robert Abelson[8] introduced **script theory**, a specialized form of schemas that encode typical event sequences in human cognition. For example, we have a "restaurant script" that allows us to know how a visit to a restaurant will unfold. This script would include typical objects, roles, and scenes for different events such as being seated, ordering, eating, and paying. Schank and Abelson used script theory to guide the development of AI systems, such as the Script Applier Mechanism (SAM), designed to understand and recall stories[12]. For example, SAM could read a news story about a man ordering food and infer that he probably paid at the end – even if that step was never mentioned. Psychological research has generally supported aspects of Schank and Abelson's script theory, e.g. people tend to recall events from a story based on their familiar order according to typical scripts, rather than the order in which the events necessarily occurred[13].

Around the same time, Rumelhart[14] proposed that schemas are "hierarchical knowledge structures" that organize all levels of understanding, allowing for generalization and inference. According to Rumelhart's conception, schemas provide "slots" or "variables" that specify the components or attributes of a given concept, and "values" or "fillers" that provide specific information for an instantiation of that concept. For instance, we may have a schema for houses that would contain variables such as "number of bedrooms", "neighbourhood", or "type of heating", which can be readily filled when encountering a new house. According to Rumlerhart's framework, new information is "integrated" into a schema when we map a given value to a given variable like this. Dedre Gentner's later work on structure-mapping theory[15] advanced a complementary view of how structured knowledge emerges through experience. Instead of treating schemas as static templates, she proposed that cognition operates through processes of relational alignment and abstraction, whereby structural correspondences between familiar and novel domains support understanding and transfer[16]. Through repeated analogical comparisons, individuals form relational schemas that generalize across contexts—a process she termed progressive alignment[17]

Combining Piaget[9]'s theories with later conceptions of schemas like those of Rumelhart[14], we can identify three distinct processes that schema theory would describe when we store or interpret new information. These three processes depend on how well new information matches existing schemas (Fig. 1):

- In cases of high fit we get **integration**; new information is rapidly and accurately stored using an existing schema.
- In cases of medium-fit we get **assimilation**; new information is rapidly stored using an existing schema, but it is modified to fit it.

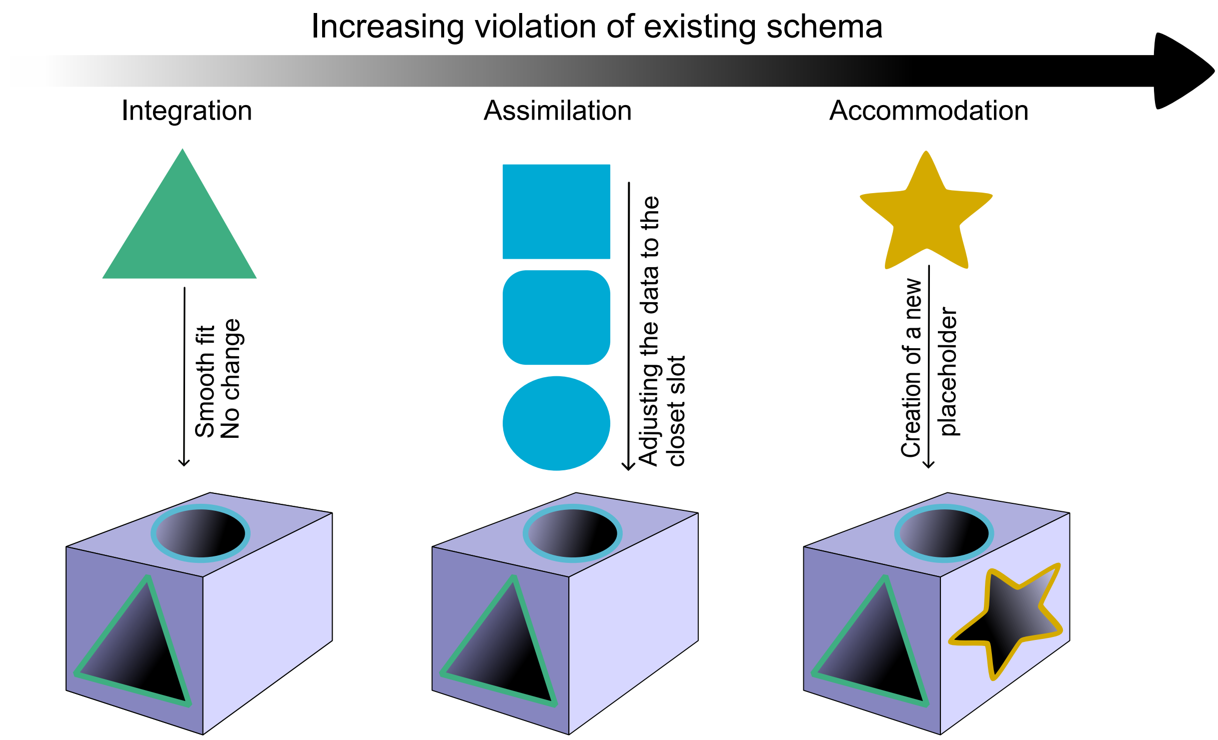

Figure 1: Schematic representation of cognitive processing based on schema theory. New information can be integrated through three main processes depending on its complexity: (1) Integration occurs when incoming data fits smoothly into an existing schema without requiring changes; (2) Assimilation requires adjusting the new data to match the closest existing schema; and (3) Accommodation involves creating a new schema when the information cannot fit into any existing structure. Complexity and the requirement for additional computation and learning increases from integration to accommodation.

- In cases of low-fit we get **accommodation**; the existing schema is updated, or a new schema is created, in order to account for the new information, requiring more time and computation.

In what follows, we will use these three processes (integration, assimilation, and accommodation) to form the foundation for our analysis of how schema theory manifests in both neuroscience and modern AI research.

Neuroscience of Schemas

Given that schemas are cognitive structures formed through repeated experiences, they are deeply intertwined with episodic memory systems[2,18]. Episodic memories themselves can be encoded and stored with varying levels of detail, a gradient that reflects anatomical differentiation along the hippocampal long axis: posterior hippocampus and associated sensory cortices support fine-grained, perceptually rich memories, while anterior hippocampus and associative cortical areas are implicated in more schematized, gist-like representations[19,20]. This continuum of schematization highlights the difficulty of clearly distinguishing a highly abstracted episodic memory from a schema proper. Schemas are further supported by prefrontal cortex (PFC) systems that integrate across episodes to extract common structure. PFC-mediated top-down control modulates both the encoding and retrieval of new information based on schema fit, the degree to which new inputs align with existing knowledge frameworks[21].

When incoming information aligns well with existing schemas, evidence suggests that the PFC exhibits enhanced activity and strong functional coupling with multiple cortical regions[6,22–25]. This interaction facilitates rapid integration of new information and may involve local synaptic plasticity within PFC circuits[26–28]. Through this interplay between the PFC and the rest of the brain, schema-consistent information can be incorporated directly into cortical representations, reducing reliance on encoding detailed episodic memory and bypassing prolonged hippocampal consolidation[5,23,27,29,30].

In cases of assimilation, when new inputs partially align with existing schemas, encoding is still facilitated but often biased. Incoming details are adjusted to conform to prior expectations, leading to distortions and less accurate recall[31]. More accurate recall of schema-inconsistent details appears to require enhanced PFC coupling with temporal and parietal regions to to overcome the incongruency and store the memory accurately[30,32,33]. Interestingly, interfering with ongoing PFC processing can sometimes prevent such distortions of information during assimilation[34].

In contrast, accommodation, which involves learning highly novel, schema-inconsistent or unrelated information, strongly engages the hippocampus and medial temporal lobe (MTL) systems[27,35–37]. The engagement of the MTL likely reflects interactions between the hippocampus and neuromodulatory systems in the midbrain and brainstem, most notably dopaminergic and adrenergic pathways[38,39]. Exposure to novel information may trigger a neuromodulatory state that promotes the formation of entirely new episodic memories in the hippocampus and related structures[39]. Over a prolonged consolidation period, and the formation of multiple novel episodic memories, a new schema can be formed or existing schemas can be updated to accommodate this new information[5,40]. Paradoxically, this process may enable individuals to remember wholly novel or highly schema-inconsistent information more effectively than slightly schema-consistent information[31,38,41].

Together, these schema-guided processes of memory formation may support the construction and maintenance of abstract, compositional knowledge structures that can enable flexible behavior and planning. Such organization of experience allows the brain to reuse and recombine learned components when facing new situations, a principle that closely mirrors hierarchical strategies in reinforcement learning[42–46]. Understanding these correspondences opens the door for computational and AI systems to integrate schema-like mechanisms, potentially enhancing data efficiency, transfer, and adaptive decision-making.

Computational Models of Schemas

In early symbolic AI, schemas were formalized as frame-based structures, with clearly defined slots and default values to be filled with contextual information, allowing for flexible yet structured reasoning[47]. This concept was expanded on to include the structure of events with “scripts”, i.e. frame-like structures that provided the scaffolding that involved a stereotyped sequence of actions or events in a particular context, such as going to a restaurant or visiting a doctor[8]. While these symbolic representations were highly influential in cognitive science and early AI, they ultimately struggled to scale to the complexity and variability of real-world inputs.

Like symbolic AI practitioners, connectionist researchers also placed emphasis on schemas. But, they tended to take a different tact. David Rumelhart very specifically highlighted schemas as critical to understanding human cognition, arguing that knowledge is organized into flexible, interconnected mental frameworks via schemas[3,48,49]. His work emphasized that the frameworks that schemas provide are crucial for making sense of the world, understanding language, remembering experiences, and acquiring new knowledge. But, this early work was largely couched at higher-level.

Later connectionist models from the 1980s, inspired by Rumelhart, re-imagined schemas as emergent phenomena in distributed neural systems. Most notably, Harmony Theory proposed that sub-symbolic cognitive systems could represent knowledge through patterns of activation across interconnected units. According to this perspective, schemas emerge as stable attractor states in a neural network's dynamics[50]. In this approach, knowledge atoms, small units of information, are activated through spreading activation, and schemas correspond to coherent, high-"harmony" configurations that integrate these atoms.

Likewise, Adaptive Resonance Theory, provided a powerful sub-symbolic account of knowledge acquisition, and implicitly schemas, not as a result of pre-defined symbolic structures but rather emergent categories or prototypes that a network can learn through unsupervised or supervised interactions with the environment[51,52].

Bridging symbolic and connectionist paradigms, Gary Drescher's constructivist approach[53] offered a computational realization of Piaget's notions of assimilation and accommodation[54], proposing that schemas can be incrementally constructed through sensorimotor experience and internal simulation. Drescher's work proposed how abstract knowledge structures could arise from low-level interactions with the environment, a perspective that presaged later cognitive accounts of self-organizing schema representations[53].

These classical works laid the foundation for later conceptualizations of schemas that built on the neuroscience of schematic memory processing. Complementary Learning Systems (CLS) theory, proposed that learning about the patterns in episodes of experience requires an interaction between a fast-learning episodic system (i.e. the hippocampus) and a slow-learning neural network that stored semantic knowledge via gradual integration of information (i.e. the neocortex)[55]. CLS was later updated to account for people and animal's abilities to weight different experiences depending on their goals, and to incorporate information into neocortical traces rapidly through schematic integration[56]. These ideas provided the background for other more recent approaches in AI that seek to address problems of continual learning more broadly[57].

Schemas are fundamentally about how past experience informs the way we store new information and turn it into knowledge. It is therefore also worth noting other computational approaches that model knowledge explicitly via graph-like structures that encode probabilistic relationships among events, objects, and agents, often learned from data using neural-symbolic methods or graph neural networks[58–61]. Such hybridization of symbolic, sub-symbolic, and graph-based approaches reflects an attempt to marry the insights from the older symbolic approaches to schemas with the more flexible, data-driven approaches provided by connectionist models[62].

However, the field of AI has moved at a very fast pace, and many of the earlier ideas about how to incorporate insights regarding memory storage have been superseded by developments in large, pretrained models. We argue that these advances invite a reconsideration of schemas not as discrete cognitive constructs, but as an emergent principle of abstraction in neural systems, aligning closely with the vision of early connectionist theories[50,51].

Evidence for Schemas in Modern AI Systems

Large-scale, modern AI models, such as large language language models (LLMs), are not architecturally designed to represent schemas explicitly. Rather, they are designed to be flexible sequence processing models, with the ability to identify how different contexts impact the interpretation of items in a sequence[63,64]. Thus, these models are very different from both early symbolic

AI models and more recent graph-based models that use explicit, frame-based or graph-based representations to structure and store new knowledge[12,47,58]. Moreover, unlike some of the early connectionist systems[50,65], LLMs are not generally framed as models of schematic knowledge representation. Yet, LLMs often display many behaviours that are reminiscent of the functional hallmarks of schema-driven learning, i.e. integration, assimilation, and accommodation.

When new information provided to an LLM aligns well with the data encountered during pre-training, LLMs can rapidly learn the new information, even without any updates to their synaptic weight parameters[66,67]. This process is often referred to as “in-context-learning”, and it results from changes to the internal activation vectors that mimic the process of learning, but with no parameter changes[68,69]. Evidence suggests that in-context learning is a result of the models using previously learned statistical and semantic relationships to interpret new data[70,71], and this can even be framed as a form of implicit Bayesian inference[72,73]. Notably, in-context learning is similar to integration via schemas as observed in humans, because it provides LLMs with the ability to encode new information using existing knowledge when they are commensurate. As well, similar to schema-based learning, in-context learning does not work when new information conflicts with previously learned information[74]. This process, sometimes referred to as “knowledge conflict”, can lead to “confirmation bias”, wherein the model only keeps those parts of the new information that match its previous data[75], or “hallucinations”, where the model invents new information that better matches its previous data[76]. Arguably, the adaptation of new information to match existing knowledge is similar to the process of assimilation in schema-based learning in humans.

When in-context learning is not sufficient, training modern LLMs on new information involves actual changes to the model’s synaptic weight parameters, a process known as “fine-tuning”[67,77,78]. Similar to accommodation, fine-tuning can be used to modify existing stored information in order to create scaffolds for new behaviours. But, as with accommodation in humans, fine-tuning can modify existing knowledge or even overwrite it[79].

Altogether, these observations suggest that LLMs have something akin to schemas, even if they are not storing information in the same way that our brains do. Moreover, these schema-like behaviours in LLMs echo earlier connectionist accounts, such as Harmony Theory, which proposed that schemas could emerge naturally in sub-symbolic systems as high-probability activation states[50]. Indeed, given that LLMs are pretrained to predict high-probability words in sentences, it is perhaps not surprising that their internal systems develop something like the schemas of Harmony theory. Another way of phrasing this is that in LLMs schemas exist and they correspond to activation patterns in the neural network that map to high-probability sequences.

To summarize, large-scale AI models are not built with schemas by design; they have neither explicit frame-like nor graph-like structures for storing information. Yet, these models exhibit the hallmarks of schemas:

- When new information matches pre-existing knowledge they can rapidly and accurately store the new information (integration)
- When new information only partially matches existing knowledge they can rapidly store new info, but they often distort it (assimilation)
- When new information doesn’t match existing knowledge, they require more extensive training to learn it, and this changes how they process future information (accommodation)

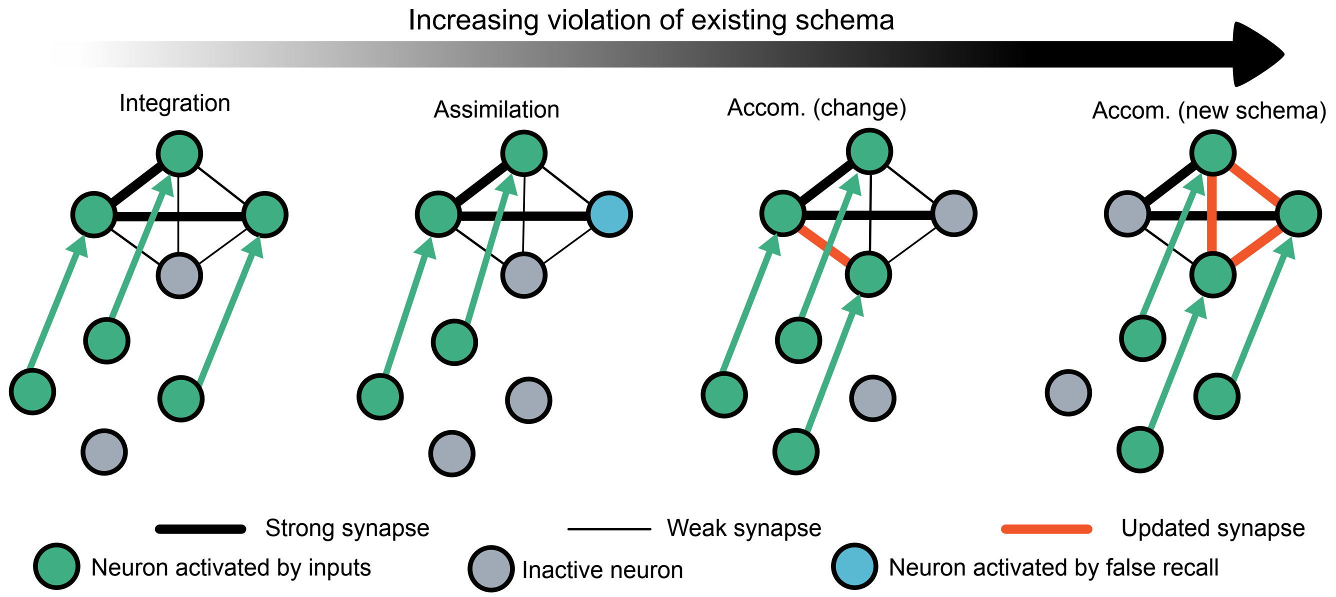

Figure 2: Schema theory mechanisms in neural networks. When new inputs are fully consistent with existing high-probability patterns (integration), activation flows along already strong synapses without significant change. With minor inconsistencies relative to existing patterns (assimilation), inputs can still be mapped onto existing connections, though false recall may occur. Greater mismatches require accommodation (change), where specific synaptic weights are updated (shown in orange) to incorporate the new information. When violations of existing high-probability patterns are too extensive (accommodation), the system engages in more widespread updating of synaptic connections to support a reorganized representation.

Lessons to be drawn in Neuroscience

What does the presence of schema-like information storage in LLMs tell us? The apparent presence of schema-like effects in modern AI models with no explicit, built-in schemas suggests the original ideas of Harmony Theory were closer to the mark than other theories of schematic storage[50]. It is not unreasonable, then, to take a few lessons from the conception of schemas that Harmony Theory initially proposed (Fig. 2):

- When new information is well-aligned to existing high-probability activation patterns (based on strong synaptic connections) learning can occur with almost no changes to synapses (integration).
- When new information partially aligns with existing high-probability activation patterns learning can occur with few changes to synapses, but this could distort the information that is recalled (assimilation).
- When new information doesn't match existing high-probability activation patterns learning requires large updates to synapses, potentially involving changes to existing high-probability patterns or construction of new high-probability activation patterns altogether (accommodation).

If we take this conception of schemas into account, then schemas are probably not specific structures in the brain, but simply high-probability activation patterns in the *abstract* latent space of the brain's neural networks. As such, the distinction between "schematized" and "detailed" memories likely just relates to how abstract the information encoded by the relevant activation patterns is. Data showing that schematic processing depends on the anterior axis of the medial temporal lobes and the PFC may simply reflect the fact that these brain regions tend to be concerned with more abstract representations. Moreover, the data showing that the hippocampus is particularly important when new information clashes with existing schemas may reflect the hippocampus' capacity for large amounts of rapid synaptic plasticity[80], and the requirement for the brain to update synapses when new information doesn't match existing high-probability activation patterns. The importance of neuromodulators in this process likely reflects both their role in detecting novelty, and potentially, the importance of reinforcement learning for shaping activation patterns in the brain[42].

Given these considerations, we suggest two key implications for neuroscience. The first is that "schemas" are conceptual tools that we scientists use to better understand processes in the brain that are otherwise hard to describe with language. Put another way, there may not be actual schemas in the brain, per se, but rather, different configurations of distributed representations that can be more or less probable depending on how past experience with higher-order patterns in our lives shaped our synaptic connectivity. The second is that there is likely a spectrum of "schematization" in the brain, related to abstraction. In other words, there is no hard dividing line between "schematized" and "non-schematized" memories, since schemas are merely high-probability activation patterns in abstract representational space. High-probability patterns can also exist in less abstract, more sensory representational space. Indeed, even low-level sensory information can be shaped by previous experience, depending on how well it matches that previous experience[81]. But, it is simply a quirk of scientific nomenclature and the history of schema theory (which is rooted in more abstract psychological studies) that we don't consider these lower-level sensory impacts to be a reflection of "schemas". Moreover, there is probably a continuous spectrum between detailed, sensory related activation patterns and more abstract, schematic activation patterns, which maps to the posterior-to-anterior axis in the human brain. This implies that episodic memories and schematized memories may not be wholly distinguishable. In-line with this, an interesting shift in the memory consolidation literature in recent years has been a growing recognition that there is not a simple relationship whereby new, episodic memories exist in the hippocampus and old, schematized memories exist in the neocortex[35]. Rather, when new information is stored, the role of the hippocampus likely has more to do with whether the new information matches existing synaptic connections. If it does, then learning can proceed rapidly with few changes to the networks in the brain, and thereby, less need for hippocampal involvement at encoding. According to this perspective, the process of "schematization" of memories[5,22,82] may actually have less to do with "transferring" memories to the neocortex, and more to do with learning abstract relationships in stored data by building high-probability activation patterns related to these relationships.

Altogether, this perspective invites a shift in how we conceptualize and discuss "schemas" within neuroscience. It suggests that we should be cautious of any model whereby schemas are an explicit, distinct knowledge structure in the brain. Instead, it brings us to a perspective, closer to Harmony Theory, which views schemas as a conceptual tool that we scientists apply to help us account for the ways in which previous, abstract knowledge stored in a neural network can impact the storage of new information.

Conclusion

These considerations suggest two sets of practical implications, one for neuroscience and one for AI design. For neuroscience, if schemas are better understood as high-probability activation patterns in distributed representations, rather than discrete symbolic structures, experimental work should focus on population-level activity patterns that capture abstract patterns, rather than any specific brain regions or cells that encode schemas[83]. Notably, this shift aligns with recent analyses using representational similarity and manifold geometry to study memory coding[35,84]. Moreover, the notion of a continuum of schematization implies that episodic and schematic memories are not categorically distinct but rather vary along gradients of abstraction, possibly mapping onto the posterior-to-anterior cortical axis and hippocampal–cortical interactions[85,86]. Sensory systems should also be considered part of this spectrum, since prior experience shapes even low-level perception across visual and auditory modalities[87]. This perspective re-frames memory consolidation as a process of matching new information to existing high-probability structures, rather than a simple temporal transfer from hippocampus to neocortex. Experimentalists should then seek to understand how the brain identifies when an activation pattern is low-probability, given current connections, and how that in turn could lead to greater engagement of the hippocampus.

For AI, lessons from schema theory help explain the recent shift away from explicitly defined symbolic schemas toward architectures in which schemas emerge implicitly as distributed probabilistic priors. This transition is exemplified by stochastic latent world models such as PlaNet[88] and Dreamer[89], which demonstrate how abstraction and generalization can arise from probabilistic dynamics in latent space rather than through hand-coded symbolic structures. Still, schema theory in neuroscience and cognitive psychology has important implications for improving the storage of new memories in AI systems. Specifically, AI systems could take inspiration from the brain and engage in a post encoding process of "schematization"[1], whereby more and more abstract patterns get identified in new data through a process of replay and consolidation that occurs using internally generated "offline" activity[90]. As these schematization become more elaborate, they support faster learning and more effective recall of information [2].

Such mechanisms could help AI systems to learn continuously from new experiences while forming meaningful semantic connections across the existing ones. Altogether, neuroscientists and AI researchers can and should continue to rely on schema theory to help guide our ideation and research, but we should avoid the temptation to think of schemas as distinct, explicit structures in either natural or artificial brains.

Acknowledgement

We acknowledge support by NSERC (Discovery Grant: RGPIN-2020-05105; Discovery Accelerator Supplement: RGPAS-2020-00031) and CIFAR (Canada AI Chair; Learning in Machine and Brains Fellowship).

[1]The "post-encoding process of schematization" refers to what the brain does after it initially forms a memory: it reorganizes that memory so it fits into broader, more abstract knowledge structures (schemas). This is not just storing the memory—it's transforming it. In AI, this process can be seen as learning adaptive abstractions.

[2]"Internally generated offline activity" refers to neural or computational activity that happens when a system is not interacting with the external environment, but is instead replaying, simulating, or reorganizing information internally.

References

1. Piaget, J. (1926). The Language and Thought of the Child. Harcourt, Brace.
2. Bartlett, F.C. (1932). Remembering: A Study in Experimental and Social Psychology. Cambridge, UK: Cambridge University Press.
3. Rumelhart, D.E. (1980). Schemata: The building blocks of cognition. In R.J. Spiro, B.C. Bruce, and W.F. Brewer, eds. Theoretical Issues in Reading Comprehension pp. 33–58.. Lawrence Erlbaum pp. 33–58.
4. Kumaran, D. (2013). Schema-driven facilitation of new hierarchy learning in the transitive inference paradigm. Neuron *79*, 735–747. doi: `10.1016/j.neuron.2013.06.019`.
5. Tse, D., Langston, R.F., Kakeyama, M., Bethus, I., Spooner, P.A., Wood, E.R., Witter, M.P., and Morris, R.G. (2007). Schemas and memory consolidation. Science *316*, 76–82.
6. van Kesteren, M.T.R., Rijpkema, M., Ruiter, D.J., and Fernández, G. (2010). Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. The Journal of Neuroscience *30*, 15888–15894. doi: `10.1523/JNEUROSCI.2674-10.2010`.
7. Minsky, M. (1975). A framework for representing knowledge. In The Psychology of Computer Vision pp. 211–277.. McGraw-Hill pp. 211–277.
8. Schank, R.C., and Abelson, R.P. (1977). Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures. Hillsdale, NJ: Lawrence Erlbaum Associates.
9. Piaget, J. (1952). The Origins of Intelligence in Children. International Universities Press.
10. Goswami, U. (2008). Principles of learning, implications for teaching: A cognitive neuroscience perspective. Journal of Philosophy of Education *42*, 381–399.
11. Schacter, D.L., Coyle, J.T., Fischbach, G.D., Moscovitch, M., and Tulving, S.E. (1999). Memory distortion: How minds, brains, and societies reconstruct the past. Philosophical Transactions of the Royal Society B: Biological Sciences *354*, 1345–1356.
12. Schank, R.C., , and Others. Sam–a story understander. research report no. 43. Tech. Rep. Yale University, Department of Computer Science New Haven, CT (1975). URL: `https://eric.ed.gov/?id=ED161024` eRIC Document Reproduction Service No. ED161024.
13. Bower, G.H., Black, J.B., and Turner, T.J. (1979). Scripts in memory for text. Cognitive psychology *11*, 177–220.
14. Rumelhart, D.E. (1980). Schemata: The building blocks of cognition. In Theoretical Issues in Reading Comprehension pp. 33–58.. Lawrence Erlbaum pp. 33–58.
15. Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. Cognitive Science *7*, 155–170. doi: `10.1207/s15516709cog0702_3`.
16. Gentner, D., Loewenstein, J., and Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. Journal of Educational Psychology *95*, 393–408.
17. Gentner, D., and Loewenstein, J. (2002). Relational language and relational thought. In D. Gentner, R.L. Goldstone, and D.L. Medin, eds. Similarity and Categorization pp. 245–277.. New York: Oxford University Press pp. 245–277.

18. Ghosh, V.E., and Gilboa, A. (2014). The schema concept: Recent advances and applications. Trends in Cognitive Sciences *18*, 504–510.

19. Poppenk, J., Evensmoen, H.R., Moscovitch, M., and Nadel, L. (2013). Long-axis specialization of the human hippocampus. Trends in Cognitive Sciences *17*, 230–240.

20. Robin, J., and Moscovitch, M. (2017). The role of the hippocampus in memory for the relationships among contexts, events, and episodes. Hippocampus *27*, 15–29.

21. van Kesteren, M.T., Rijpkema, M., Ruiter, D.J., and Fernandez, G. (2012). Schema-dependent encoding processes relate to academic performance. Journal of Cognitive Neuroscience *24*, 1332–1341.

22. Audrain, S., and McAndrews, M.P. (2022). Schemas provide a scaffold for neocortical integration of new memories over time. Nature Communications *13*, 5795. doi: `10.1038/s41467-022-33517-0`.

23. Sommer, T., Hennies, N., Lewis, P.A., and Alink, A. (2022). The assimilation of novel information into schemata and its efficient consolidation. Journal of Neuroscience *42*, 5916–5929. doi: `10.1523/JNEUROSCI.2373-21.2022`.

24. Shohamy, D., and Adcock, R.A. (2010). Dopamine and adaptive memory. Trends in Cognitive Sciences *14*, 464–472.

25. Brod, G., Lindenberger, U., Werkle-Bergner, M., and Shing, Y.L. (2015). Differences in the neural signature of remembering schema-congruent and schema-incongruent events. NeuroImage *117*, 358–366. URL: `https://www.sciencedirect.com/science/article/pii/S1053811915004760`. doi: `https://doi.org/10.1016/j.neuroimage.2015.05.086`.

26. Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., Bito, H., and Morris, R.G.M. (2011). Schema-dependent gene activation and memory encoding in neocortex. Science *333*, 891–895. doi: `10.1126/science.1205274`.

27. Bonasia, K., Sekeres, M.J., Gilboa, A., Grady, C.L., Winocur, G., and Moscovitch, M. (2018). Prior knowledge modulates the neural substrates of encoding and retrieving naturalistic events at short and long delays. Neurobiology of Learning and Memory *153*, 26–39. doi: `10.1016/j.nlm.2018.02.017`.

28. van Kesteren, M.T., Beul, S.F., Takashima, A., Henson, R.N., Ruiter, D.J., and Fernández, G. (2013). Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: From congruent to incongruent. Neuropsychologia *51*, 2352–2359. doi: `10.1016/j.neuropsychologia.2013.05.027`.

29. Sekeres, M.J., Winocur, G., and Moscovitch, M. (2018). The hippocampus and related neocortical structures in memory transformation. Neuroscience Letters *680*, 39–53. doi: `10.1016/j.neulet.2018.05.006`.

30. van der Linden, M., Berkers, R.M., Morris, R.G., and Fernández, G. (2017). Angular gyrus involvement at encoding and retrieval is associated with durable but less specific memories. Journal of Neuroscience *37*, 9474–9485. doi: `10.1523/JNEUROSCI.3603-16.2017`.

31. Greve, A., Cooper, E., Tibon, R., and Henson, R.N. (2019). Knowledge is power: Prior knowledge aids memory for both congruent and incongruent events, but in different ways. Journal of Experimental Psychology: General *148*, 325–341. doi: `10.1037/xge0000498`.

32. Guo, D., Chen, G., and Yang, J. (2023). Effects of schema on the relationship between post-encoding brain connectivity and subsequent durable memory. Scientific Reports *13*, 8736. doi: 10.1038/s41598-023-34822-4.

33. Webb, C.E., Turney, I.C., and Dennis, N.A. (2016). What's the gist? the influence of schemas on the neural correlates underlying true and false memories. Neuropsychologia *93*, 61–75. doi: 10.1016/j.neuropsychologia.2016.09.023.

34. Berkers, R.M., van der Linden, M., de Almeida, R.F., Müller, N.C., Bovy, L., Dresler, M., Morris, R.G., and Fernández, G. (2017). Transient medial prefrontal perturbation reduces false memory formation. Cortex *88*, 42–52. doi: 10.1016/j.cortex.2016.12.015.

35. Sekeres, M.J., Schomaker, J., Nadel, L., and Tse, D. (2024). To update or to create? the influence of novelty and prior knowledge on memory networks. Philosophical Transactions of the Royal Society B: Biological Sciences *379*, 20230238. doi: 10.1098/rstb.2023.0238.

36. Kumaran, D., and Maguire, E.A. (2007). Match–mismatch processes underlie human hippocampal responses to associative novelty. The Journal of Neuroscience *27*, 8517–8524. doi: 10.1523/JNEUROSCI.1677-07.2007.

37. Van Kesteren, M.T.R., Ruiter, D.J., Fernandez, G., and Henson, R.N. (2012). Schema theory revisited: The role of schemas in memory stabilization and updating. Trends in Cognitive Sciences *16*, 511–518.

38. Lisman, J.E., and Grace, A.A. (2005). The hippocampal-vta loop: controlling the entry of information into long-term memory. Neuron *46*, 703–713. doi: 10.1016/j.neuron.2005.05.002.

39. Duszkiewicz, A.J., McNamara, C.G., Takeuchi, T., and Genzel, L. (2019). Novelty and dopaminergic modulation of memory persistence: A tale of two systems. Trends in Neurosciences *42*, 102–114. doi: 10.1016/j.tins.2018.10.002.

40. Richards, B.A., Xia, F., Santoro, A., Husse, J., Woodin, M.A., Josselyn, S.A., and Frankland, P.W. (2014). Patterns across multiple memories are identified over time. Nature Neuroscience *17*, 981–986.

41. Takeuchi, T., Duszkiewicz, A.J., Sonneborn, A., Spooner, P.A., Yamasaki, M., Watanabe, M., Smith, C.C., Fernández, G., Deisseroth, K., Greene, R.W., and Morris, R.G. (2016). Locus coeruleus and dopaminergic consolidation of everyday memory. Nature *537*, 357–362. doi: 10.1038/nature19325.

42. Bein, O., and Niv, Y. (2025). Schemas, reinforcement learning and the medial prefrontal cortex. Nature Reviews Neuroscience pp. 1–17.

43. Botvinick, M.M., Niv, Y., and Barto, A.C. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. Cognition *113*, 262–280.

44. Collins, A.G., and Frank, M.J. (2013). Cognitive control over learning: Creating, clustering, and generalizing task-set structure. Psychological Review *120*, 190–229.

45. Frank, M.J., and Badre, D. (2012). Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: computational analysis. Cerebral Cortex *22*, 509–526. doi: 10.1093/cercor/bhr114.

46. Badre, D., and Nee, D.E. (2018). Frontal cortex and the hierarchical control of behavior. Trends in Cognitive Sciences *22*, 170–188.

47. Minsky, M. (1974). A framework for representing knowledge. In P.H. Winston, ed. The Psychology of Computer Vision pp. 211–277.. McGraw-Hill pp. 211–277.

48. Rumelhart, D.E., and Ortony, A. (1977). The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro, and W.E. Montague, eds. Schooling and the Acquisition of Knowledge pp. 99–135.. Hillsdale, NJ: Lawrence Erlbaum Associates pp. 99–135.

49. Rumelhart, D.E., and Norman, D.A. Accretion, tuning and restructuring: Three modes of learning. Tech. Rep. 7602 Center for Human Information Processing, University of California, San Diego La Jolla, CA (1976). URL: `https://apps.dtic.mil/sti/tr/pdf/ADA030406.pdf`.

50. Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. Parallel Distributed Processing: Explorations in the Microstructure of Cognition *1*, 194–281.

51. Carpenter, G.A., and Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. Computer Vision, Graphics, and Image Processing *37*, 54–115. doi: `10.1016/S0734-189X(87)80014-2`.

52. Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. Cognitive Science *11*, 23–63. URL: `https://www.sciencedirect.com/science/article/pii/S0364021387800253`. doi: `https://doi.org/10.1016/S0364-0213(87)80025-3`.

53. Drescher, G.L. (1991). Made-up Minds: A Constructivist Approach to Artificial Intelligence. Cambridge, MA: MIT Press.

54. Drescher, G.L. (1987). A mechanism for early piagetian learning. In AAAI. American Association for Artificial Intelligence pp. 52–56. URL: `https://cdn.aaai.org/AAAI/1987/AAAI87-052.pdf`.

55. McClelland, J.L., McNaughton, B.L., and O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models. Psychological Review *102*, 419–457.

56. Kumaran, D., Hassabis, D., and McClelland, J.L. (2016). What learning systems do intelligent agents need? complementary learning systems theory updated. Trends in Cognitive Sciences *20*, 512–534.

57. Anand, N., and Precup, D. (2023). Prediction and control in continual reinforcement learning. In Advances in Neural Information Processing Systems. URL: `https://arxiv.org/abs/2312.11669`.

58. Battaglia, P.W., Hamrick, J.B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R. et al. (2018). Relational inductive biases, deep learning, and graph networks. URL: `https://arxiv.org/abs/1806.01261`.

59. Kansky, K., Silver, T., Mély, D.A., Eldawy, M., Lázaro-Gredilla, M., Lou, X., Dorfman, N., Sidor, S., Phoenix, S., and George, D. (2017). Schema networks: Zero-shot transfer with a generative causal model of intuitive physics. In Proceedings of the 34th International Conference on Machine Learning vol. 70. PMLR pp. 1809–1818.

60. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. (2018). Graph attention networks. In A.B. Gjorgjevikj, ed. International Conference on Learning Representations, ICLR 2018. OpenReview.net. URL: https://arxiv.org/abs/1710.10903.

61. Tenenbaum, J.B., Kemp, C., Griffiths, T.L., and Goodman, N.D. (2011). How to grow a mind: Statistics, structure, and abstraction. Science *331*, 1279–1285. doi: 10.1126/science.1192788.

62. Lake, B.M., Ullman, T.D., Tenenbaum, J.B., and Gershman, S.J. (2017). Building machines that learn and think like people. Behavioral and Brain Sciences *40*, e253.

63. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, , and Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems *30*.

64. Naveed, H., Khan, A.U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., and Mian, A. (2025). A comprehensive overview of large language models. ACM Transactions on Intelligent Systems and Technology *16*, 1–39. doi: 10.1145/3744746.

65. Hinton, G.E. (1986). Learning distributed representations of concepts. In Proceedings of the Eighth Annual Conference of the Cognitive Science Society. pp. 1–12.

66. Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. et al. (2020). Language models are few-shot learners. In Advances in neural information processing systems vol. 33. pp. 1877–1901.

67. Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E.H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., and Fedus, W. (2022). Emergent abilities of large language models. . URL: https://arxiv.org/abs/2206.07682. arXiv:2206.07682.

68. Akyürek, E., Andreas, J., and Zhou, T. (2022). Learning to learn with in-context learning. In International Conference on Learning Representations.

69. von Oswald, J., Niklasson, E., Randazzo, E., Sacramento, J., Mordvintsev, A., Zhmoginov, A., and Vladymyrov, M. (2023). Transformers learn in-context by gradient descent. In Proceedings of the 40th International Conference on Machine Learning vol. 202. PMLR pp. 35151–35174. URL: https://proceedings.mlr.press/v202/von-oswald23a.html.

70. Olsson, C., Elhage, N., Nanda, N., Joseph, N., DasSarma, N., Henighan, T., Mann, B., Askell, A., Bai, Y., Chen, A. et al. (2022). In-context learning and induction heads. arXiv preprint arXiv:2209.11895.

71. Ren, J., Guo, Q., Yan, H., Liu, D., Zhang, Q., Qiu, X., and Lin, D. (2024). Identifying semantic induction heads to understand in-context learning. In Findings of the Association for Computational Linguistics: ACL 2024. Association for Computational Linguistics pp. 7135–7149. URL: https://aclanthology.org/2024.findings-acl.412/.

72. Xie, S.M., Raghunathan, A., Liang, P., and Ma, T. (2021). An explanation of in-context learning as implicit bayesian inference. In International Conference on Learning Representations.

73. Akyürek, E., Schuurmans, D., Andreas, J., Ma, T., and Zhou, D. (2022). What learning algorithm is in-context learning? investigations with linear models. arXiv preprint arXiv: 2211.15661.

74. Xu, R., Qi, Z., Guo, Z., Wang, C., Wang, H., Zhang, Y., and Xu, W. (2024). Knowledge conflicts for llms: A survey. . URL: `https://arxiv.org/abs/2403.08319`. `arXiv:2403.08319`.

75. Xie, J., Zhang, K., Chen, J., Lou, R., and Su, Y. (2023). Adaptive chameleon or stubborn sloth: Revealing the behavior of large language models in knowledge conflicts. In The Twelfth International Conference on Learning Representations.

76. Guerreiro, N.M., Alves, D.M., Waldendorf, J., Haddow, B., Birch, A., Colombo, P., and Martins, A.F. (2023). Hallucinations in large multilingual translation models. Transactions of the Association for Computational Linguistics *11*, 1500–1517.

77. Han, Z., Gao, C., Liu, J., Zhang, J., and Zhang, S.Q. (2024). Parameter-efficient fine-tuning for large models: A comprehensive survey. arXiv preprint arXiv:2403.14608.

78. Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Li, Y. et al. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223.

79. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D., Clopath, C., Kumaran, D., and Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. Proceedings of the National Academy of Sciences *114*, 3521–3526. doi: `10.1073/pnas.1611835114`.

80. Bittner, K.C., Milstein, A.D., Grienberger, C., Romani, S., and Magee, J.C. (2017). Behavioral time scale synaptic plasticity underlies ca1 place fields. Science *357*, 1033–1036. doi: `10.1126/science.aan3846`.

81. Snyder, J.S., Schwiedrzik, C.M., Vitela, A.D., and Melloni, L. (2015). How previous experience shapes perception in different sensory modalities. Frontiers in Human Neuroscience *9*, 594. URL: `https://www.frontiersin.org/articles/10.3389/fnhum.2015.00594/full`. doi: `10.3389/fnhum.2015.00594`.

82. Gilboa, A., and Marlatte, H. (2017). Neurobiology of schemas and schema-mediated memory. Trends in Cognitive Sciences *21*, 618–631.

83. Baraduc, P., Duhamel, J.R., and Wirth, S. (2019). Schema cells in the macaque hippocampus. Science *363*, 635–639.

84. Momennejad, I. (2024). Memory, space, and planning: Multiscale predictive representations. arXiv preprint arXiv:2401.09491. URL: `https://arxiv.org/abs/2401.09491`.

85. Staresina, B.P., Buzsáki, G., and Klinzing, J.G. (2024). Coupled sleep rhythms for memory consolidation. Trends in Cognitive Sciences *28*, 578–591. doi: `10.1016/j.tics.2024.04.005`.

86. Yang, S., Huang, T., Fan, R. et al. (2025). Time-dependent consolidation mechanisms of durable memory in spaced learning. Communications Biology *8*, 1234. doi: `10.1038/s42003-025-07964-6`.

87. Snyder, J.S., Schwiedrzik, C.M., Vitela, A.D., and Melloni, L. (2015). How previous experience shapes perception in different sensory modalities. Frontiers in Human Neuroscience *9*, 594. URL: `https://www.frontiersin.org/articles/10.3389/fnhum.2015.00594/full`. doi: `10.3389/fnhum.2015.00594`.

88. Hafner, D., Lillicrap, T., Fischer, I., Villegas, R., Ha, D., Lee, H., and Davidson, J. (2019). Learning latent dynamics for planning from pixels. In Proceedings of the 36th International Conference on Machine Learning (ICML). pp. 2555–2565. URL: https://arxiv.org/abs/1811.04551.

89. Hafner, D., Pasukonis, J., Ba, J., and Lillicrap, T. (2023). Mastering diverse domains through world models. Nature *623*, 756–762. doi: 10.1038/s41586-023-06881-9.

90. Kumaran, D., Hassabis, D., and McClelland, J.L. (2016). What learning systems do intelligent agents need? complementary learning systems theory updated. Trends in Cognitive Sciences *20*, 512–534. doi: 10.1016/j.tics.2016.05.004.