

Kognitionspsychologie: Session 7

Memory

Rui Mata, HS2025

Version: November 18, 2025

Learning Objectives

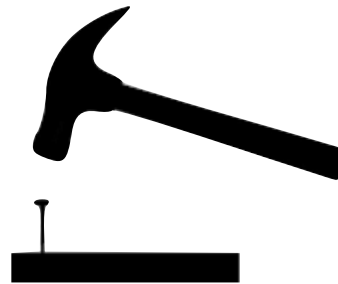
- Be able to discuss different types of human knowledge and the cognitive (memory) systems that implement it...
- Discuss the **adaptive significance** of different types of knowledge representations
- Discuss advantages and limits of **comparative approaches** to understand knowledge representation
- Be aware of general **developmental patterns** in the acquisition of knowledge
- Be able to identify central features of **cognitive and neural model(s)** of semantic cognition

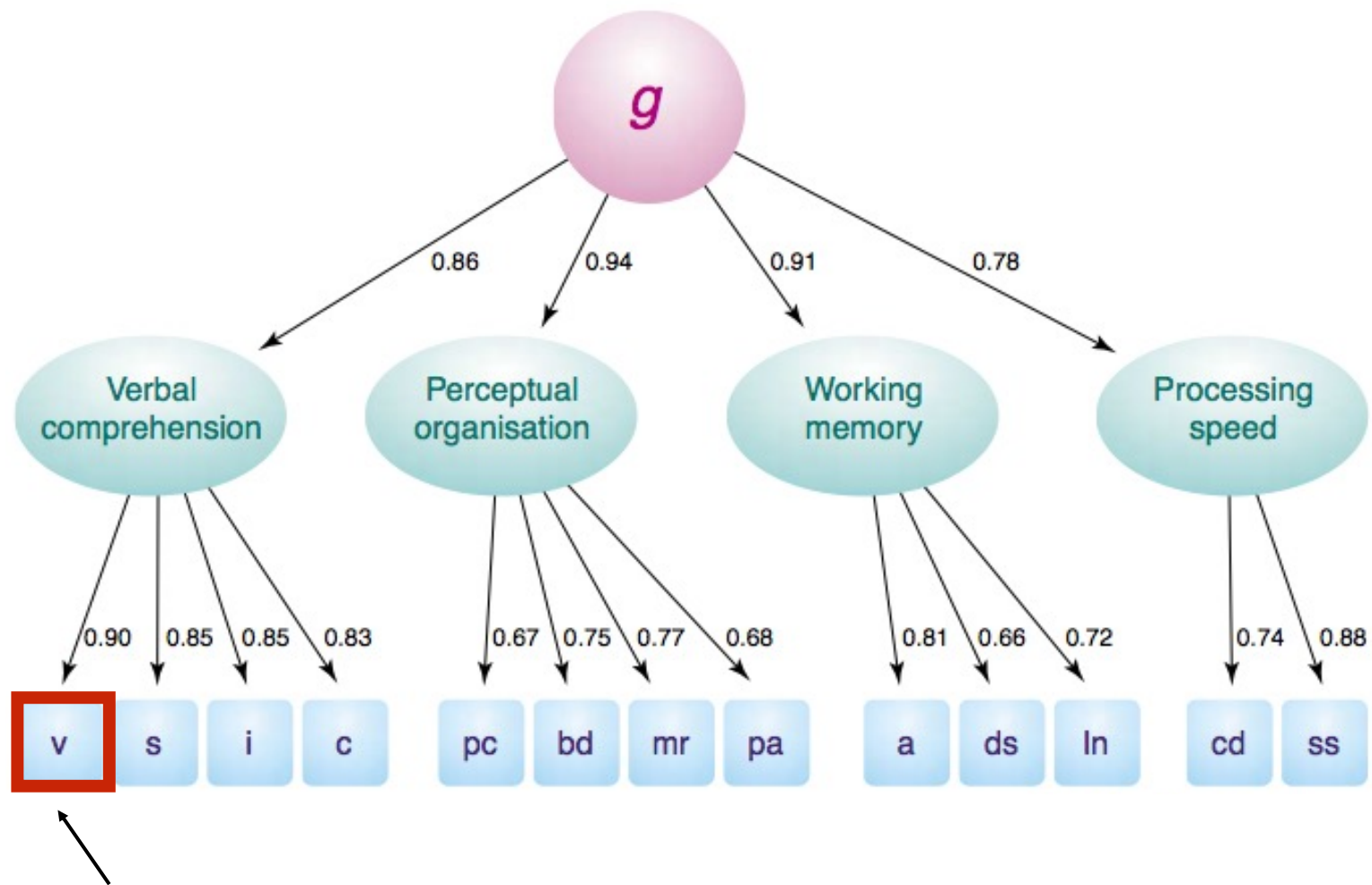
Where did you park your bike/car?

What is the capital of France?

Semantics

Ontogeny	Mechanism
Phylogeny	Adaptive Significance



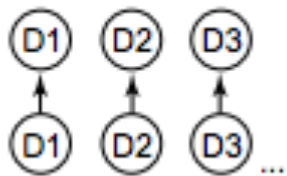
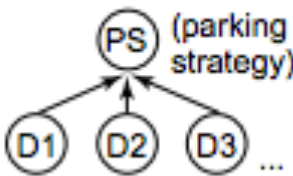


Declarative, verbal knowledge (as measured by a vocabulary test)

Deary, I. J. (2001). Human intelligence differences: A recent history. *Trends in Cognitive Sciences*, 5(3), 127–130.

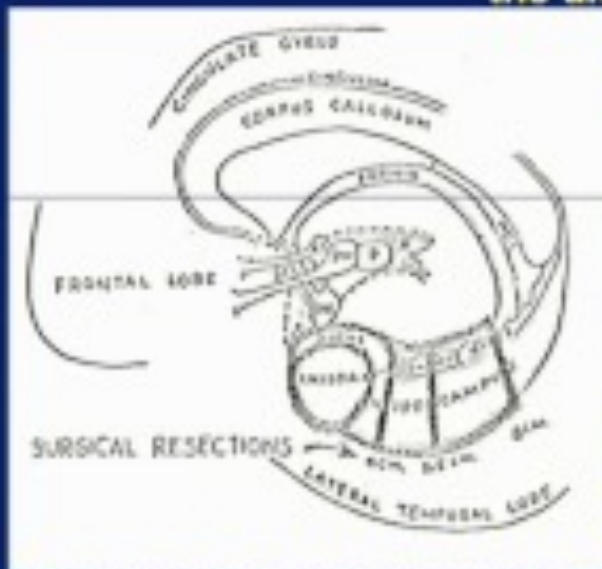
Different types of knowledge systems fulfil different goals

“The complementary learning systems framework is based on the logic of tradeoffs between mutually incompatible computational goals. The central tradeoff behind our framework involves two basic types of learning that an organism must engage in – learning about specifics versus generalities – which require conflicting neural architectures”

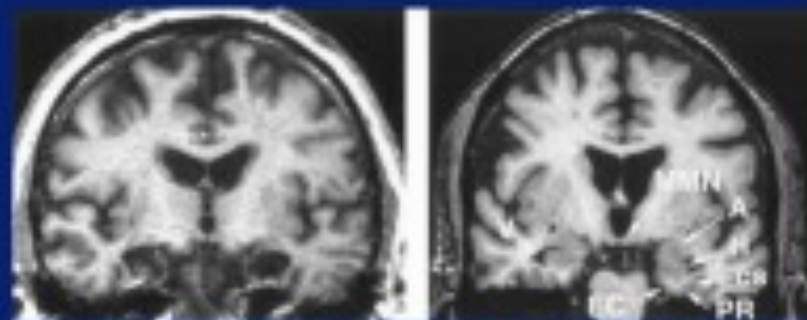
Two incompatible goals		Remember specifics	Extract generalities
Example:		<i>Where is car parked?</i>	<i>Best parking strategy?</i>
Need to:		Avoid interference	Accumulate experience
Solution:	(1)	Separate representations (keep days separate)	Overlapping representations (integrate over days)
			
	(2)	Fast learning (encode immediately)	Slow learning (integrate over days)
	(3)	Learn automatically (encode everything)	Task driven learning (extract relevant stuff)
System:		Hippocampus	Neocortex

THE AMNESIC PATIENT H.M.

- 1926** Birth
- 1942** Age 16, First major seizure
- 1953** Age 27, Bilateral medial temporal-lobe resection
- 1955** Report of pervasive and profound anterograde amnesia by Dr. Brenda Milner
- 1962** Neuropsychological examinations characterizing the amnesic syndrome



© source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>.

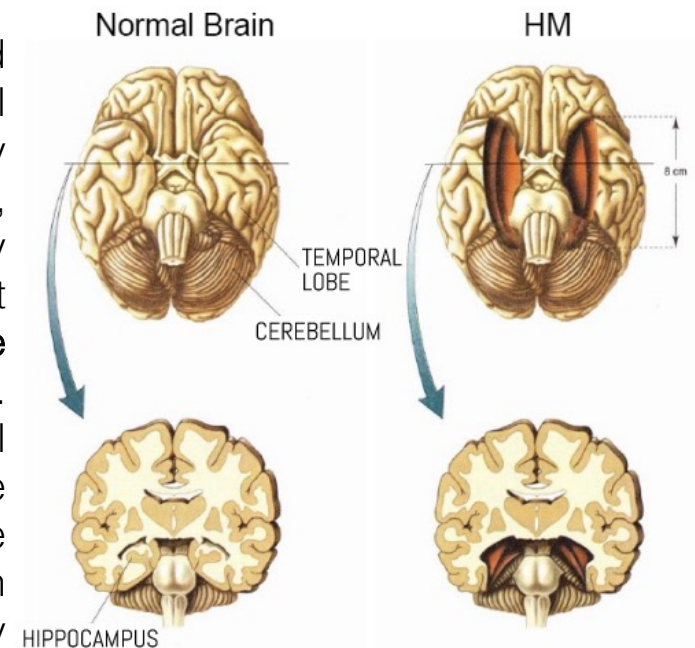


Source: Corkin, S., et al. *J Neurosci* 17, no. 10 (1997): 3964-79. © The Journal of Neuroscience. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>.

Hippocampus and entorhinal cortex (H and EC) in normal subject (left), but absent in patient H. M. (right)

A classic: H.M.

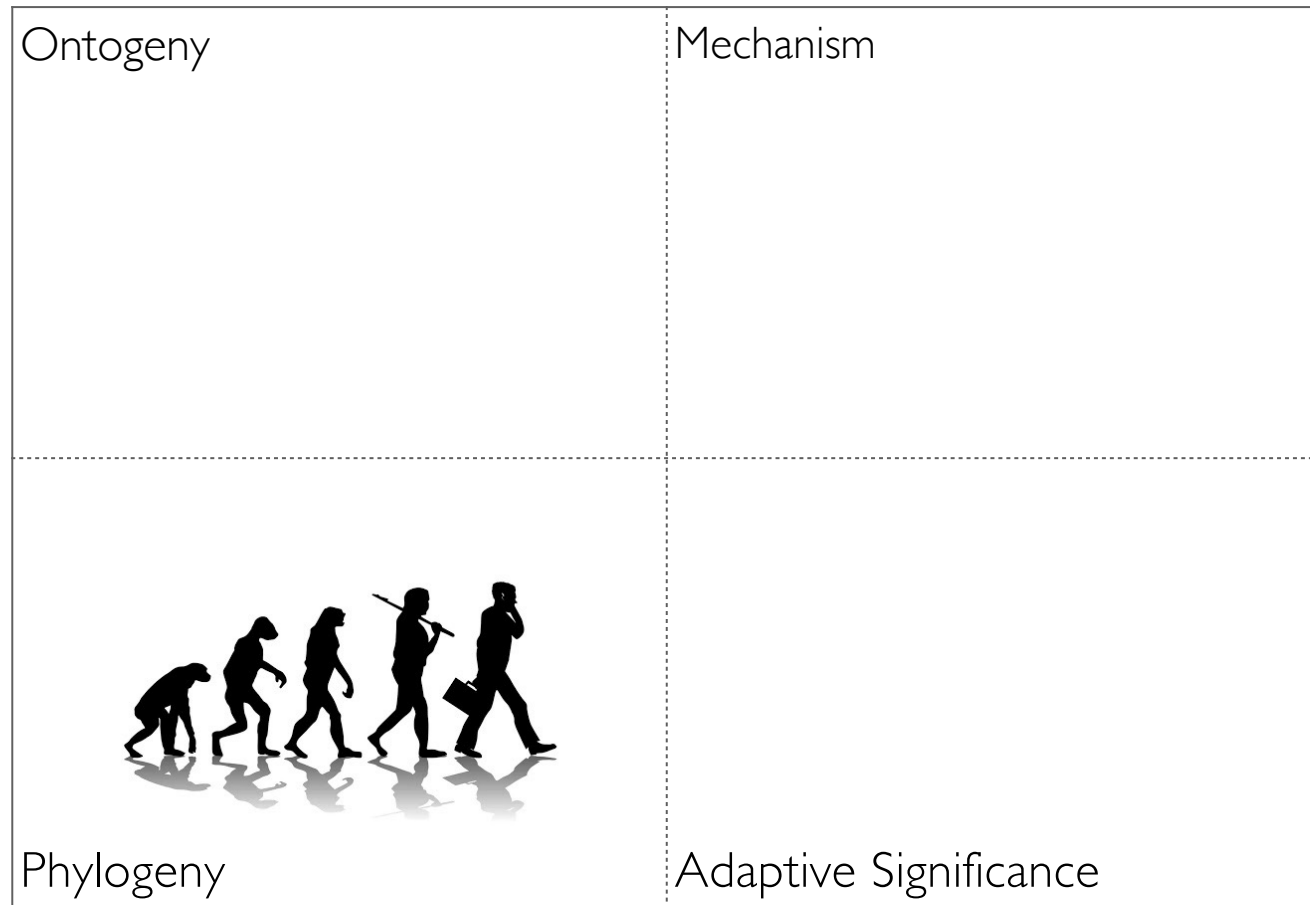
H.M. (Henry Molaison) had a history of epileptic seizures and underwent a surgical intervention that involved resecting the medial aspect of the temporal lobe bilaterally. The lesion was bilaterally symmetrical and included large portions of the temporal cortex, including the hippocampus and surrounding structures. The surgery successfully reduced the frequency and severity of his seizures but left H.M. with memory deficits, characterized by severe **anterograde amnesia**, making him unable to form new declarative memories. Additionally, he exhibited **retrograde amnesia** with a temporal gradient, losing memories of events from the years preceding the surgery while retaining older, more consolidated memories. Despite these deficits, H.M. retained the ability to learn procedural skills, such as mirror-tracing tasks, demonstrating that non-declarative memory relies on brain regions outside the medial temporal lobe.



A few conclusions can be drawn from H.M.'s case (and similar ones):

1. Memory is a distinct cerebral ability that is separate from other cognitive functions, such as perception, personality, or motivation.
2. Short-term memory and long-term memory are distinct functions: H.M. had severely impaired long-term memory, however, he could maintain and use information for a short time in immediate memory (working memory) so long as the material could be effectively rehearsed. With distraction, the information was lost.
3. Medial temporal lobe structures are not the ultimate repository of long-term memory because H.M.'s memory for remote events remained largely intact.

Semantics



Opportunities and challenges of comparative approaches

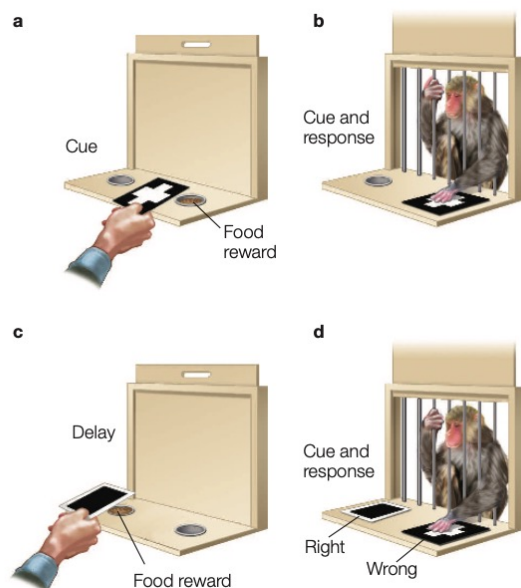


FIGURE 9.19 Delayed nonmatch-to-sample task.

(a) The correct response has a food reward located under it. (b) The monkey is shown the correct response, which will yield a reward for the monkey. (c) The door is closed, and the reward is placed under a second response option. (d) The monkey is then shown two options and must pick the correct response (the one that does not match the original sample item) to get the reward. Here the monkey is pictured making an error.

Table 1 Effects of removal of amygdala and hippocampus on memory											
Groups		Preoperative		Postoperative		Delays (% correct)			Objects (% correct)		
		Trials	Errors	Trials	Errors	30 s	60 s	120 s	3	5	10
Normal control	1	100	26	0	0	97	98	96	96	93	91
	2	80	28	0	0	99	100	98	97	97	94
	3	40	19	0	0	98	99	98	97	96	92
Amygdectomy	1	120	42	80	32	95	95	95	91	92	82
	2	100	27	340	85	91	89	92	88	87	77
	3	80	30	0	0	91	95	94	96	95	87
Hippocampectomy	1	60	17	80	22	98	93	94	95	92	84
	2	100	26	120	32	85	89	83	89	85	71
	3	120	31	20	4	98	99	95	95	92	88
Amygdectomy + hippocampectomy	1	210	49	760	179	79	65	65	62	64	59
	2	100	26	1,500	429*	64	59	63	60	55	61
	3	80	22	700	203	61	47	52	53	58	44
Group means											
Normal control		73	24	0	0	98	99	98	97	96	92
Amygdectomy		100	33	140	39	94	93	94	92	91	82
Hippocampectomy		93	25	73	19	94	94	91	93	90	81
Amygdectomy + hippocampectomy		130	32	987	270	68	57	60	58	59	55

Scores in preoperative and postoperative columns are the numbers of trials and errors preceding criterion of 90 correct choices of the novel object in 100 trials (delay following familiarisation with the other object in the pairs was 10 s). Scores in delays columns are percentage correct in 100 trials at each of three longer delays tested in succession at the rate of 20 trials per day, except for the longest delay (120 s) which was tested for 10 trials per day. Scores in objects columns are percentage correct in 150 trials for each of three multiple-object conditions tested in succession: in the first condition, three objects were presented for familiarisation, one at a time at 20 s intervals, and then re-presented in the same order, each being paired with a novel object, again at 20 s intervals; in the next condition, five objects were presented one at a time at 20 s intervals, and so on. Thirty trials were presented daily, ten sets of 3's or six sets of 5's or three sets of 10's. The minimum delay between familiarisation and choice was 60, 100, and 200 s for the three conditions, respectively. Scores for individual animals are shown in the upper part of the table, group means in the lower part. Histological examination indicated that the lesions were as intended except in animals 'hippocampectomy 2' and 'amygdectomy + hippocampectomy no. 2', both of which sustained, in addition to the planned removals, bilaterally asymmetrical damage to the ventral part of inferior temporal ('visual') cortex.

*Failed; final score, 85 correct in 100 trials.

Mishkin (1978) examined the effects of hippocampal (and amygdala) lesions on memory in monkeys. It found that while hippocampal damage alone caused mild memory impairments, the combined removal of the hippocampus and amygdala resulted in severe deficits in recognition and associative memory tasks. Later it became clear that it was not the amygdala lesion per se but the loss of surrounding tissue – parahippocampal regions – that led to deficits. The findings highlight the role of the hippocampus and parahippocampal regions as key parts of a memory network.

Mishkin, M. (1978). Memory in monkeys severely impaired by combined but not by separate removal of amygdala and hippocampus. *Nature*, 273(5660), 297–298. <https://doi.org/10.1038/273297a0>

Opportunities and challenges of comparative approaches

Table 2

Performance of Amnesic Patients and Monkeys With H⁺A⁺ Lesions on the Same Tasks

Test	Amnesic patients		Monkeys with H ⁺ A ⁺ lesions	
	+/-	Reference	+/-	Reference
Delayed nonmatching to sample	+	Squire, Zola-Morgan, & Chen, 1988; Oscar-Berman & Bonner, 1985	+	Mishkin, 1978; Zola-Morgan & Squire, 1985
Retention of object discrimination	+	Squire, Zola-Morgan, & Chen, 1988	+	Zola-Morgan & Squire, 1985
8-pair concurrent discrimination	+	Squire et al., 1988; Oscar-Berman & Bonner, 1985	+	Zola-Morgan & Squire, 1985
Object reward association	+	Squire et al., 1988	+	Phillips & Mishkin, 1984
24-hr concurrent discrimination	+	Squire et al., 1988	-	Malamut, Saunders, & Mishkin, 1984
Motor skill learning	-	Pursuit rotor task; Brooks & Baddeley, 1976	-	Lifesaver task; Zola-Morgan & Squire, 1984
Pattern discrimination	+	Predicted outcome; not yet tested	-	Zola-Morgan & Squire, 1984

Note. References are to representative studies and are not exhaustive. Plus sign indicates impairment; minus sign indicates no impairment. Monkeys may approach the 24-hr concurrent-discrimination task and the pattern-discrimination task differently than humans approach these two tasks. Humans try simply to memorize which stimulus is correct and which is incorrect (i.e., using declarative memory). Monkeys gradually learn incrementally, perhaps by gradually strengthening associations or by “tuning in” relevant dimensions of the stimuli (for fuller discussion, see Zola-Morgan & Squire, 1984). From “Neuropsychological Investigations of Memory and Amnesia: Findings From Humans and Nonhuman Primates” (p. 446) by S. Zola-Morgan and L. R. Squire, 1990, in A. Diamond, *The Development and Neural Bases of Higher Cognitive Functions*, New York: New York Academy of Sciences. Copyright 1990 by the New York Academy of Sciences. Adapted by permission.

The table above suggests strong similarities between amnesic patients and monkeys with hippocampal and amygdala lesions. These and other findings suggest that the organization and function of the medial temporal lobe is highly conserved across species (e.g., monkey, human). However, because of the use of different strategies (e.g., monkeys learn visual discrimination through habit learning rather than memorization), it took decades to uncover many of similarities in the role of these structures for memory.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99, 2, 195-231. <http://doi.org/10.1037/0033-295x.99.2.195>

Clark, R. E., & Squire, L. R. (2013). Similarity in form and function of the hippocampus in rodents, monkeys, and humans. *Proceedings of the National Academy of Sciences*, 110, 10365–10370. <http://doi.org/10.1073/pnas.1301225110>

An overview of memory systems

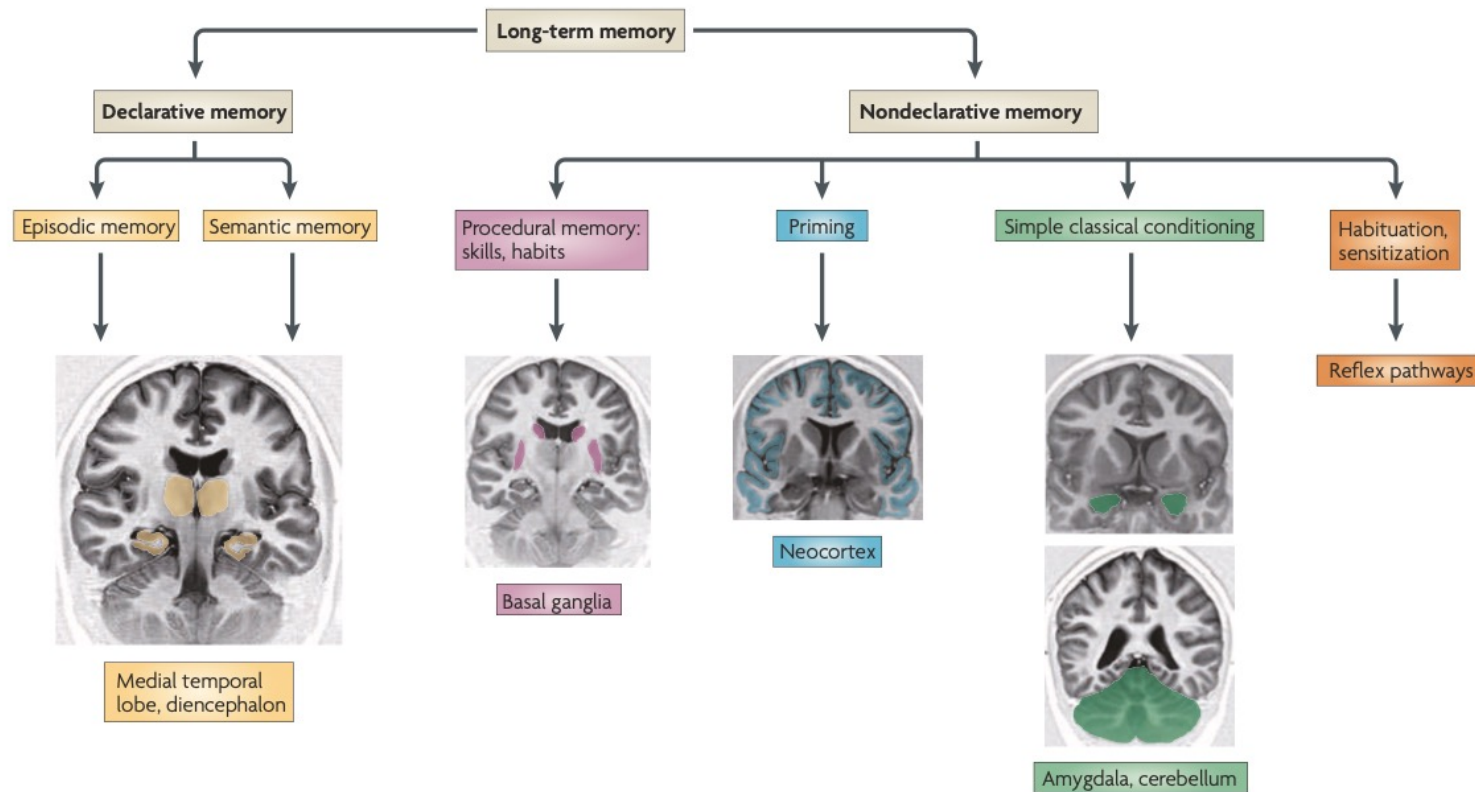



Figure 1 | **The declarative versus nondeclarative memory account.** In this model¹³, long-term memory is divided into two broad classes. One class is characterized by the capacity for conscious recollection and is called declarative memory. The other class encompasses diverse unconscious learning and memory abilities and is referred to as nondeclarative memory. Declarative memory consists of two subclasses: episodic memory, which consists of memories for autobiographical events; and semantic memory, which consists of facts and general knowledge^{34,35,143}. Semantic memories are impersonal and devoid of autobiographical context, whereas episodic

memories are personal. These include where and when episodes happened and are accompanied by a feeling of retrieving personally experienced episodes (autonoetic consciousness^{34,35,143}). In this model both episodic and semantic memories are dependent on the medial temporal lobe and diencephalon. Nondeclarative memory in this model includes procedural learning of sensorimotor and cognitive skills and habits, priming, simple conditioning, and habituation and sensitization (BOX 1), all of which are expressed in behavioural changes and are independent of the medial temporal lobe.

The combined strands of evidence contributed to a general model of human memory that distinguished different types of memory (note this model is not without limitations and remains a helpful simplification).

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience*, 11(7), 523–532. <https://doi.org/10.1038/nrn2850>

Semantics

Ontogeny	Mechanism
	
Phylogeny	Adaptive Significance

Acquiring Semantic Representations

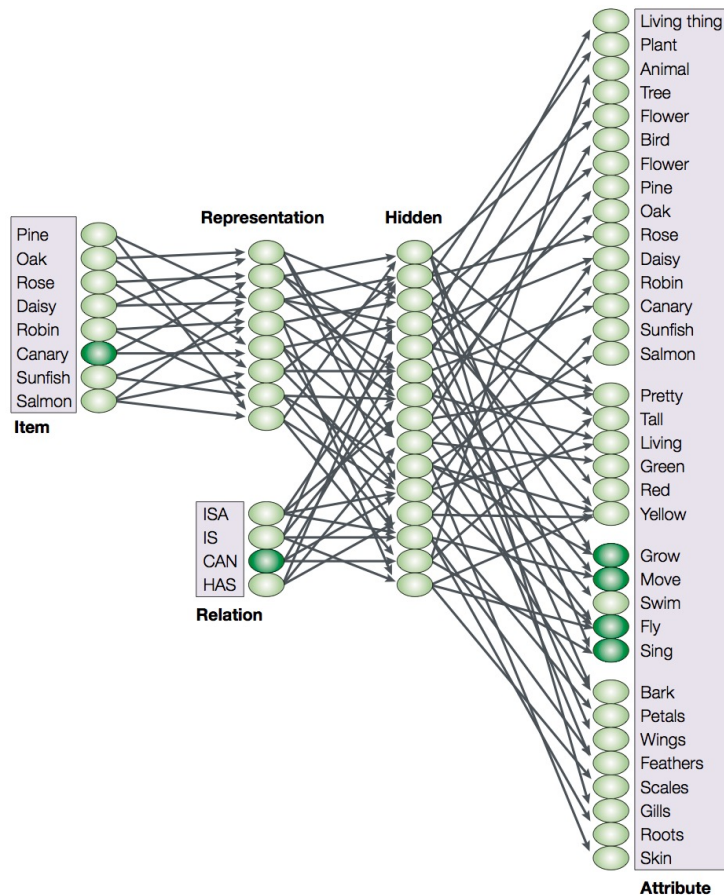


“it is well-established that early categorization abilities become refined over the developmental trajectory. Researchers have identified a global-to-basic shift in early categorical thinking, such that preverbal infants discriminate between global-level categories (i.e., dogs, cats, chairs, tables, etc.) before basic-level categories (i.e., different breeds of cats and dogs). (...) There is evidence to suggest that infants also use dynamic, causal, and functional information to guide their object categorization and discrimination.”

Poulin-Dubois, D. & Pauen, S. (2017). The development of object categories: What, when, and how? In: H. Cohen & C. Lefebvre (Eds.), *Handbook of Categorization in Cognitive Science* (2nd Ed., pp. 653–708). Elsevier.
doi:10.1016/b978-0-08-101107-2.00027-0

Acquiring Semantic Representations

Distributed Networks



The network is used to simulate learning propositions about concepts. The entire set of units used in the network is shown. Inputs are presented on the left, and activation propagates from left to right. Where connections are indicated, every unit in the pool on the left (sending) side projects to every unit on the right (receiving) side. An input consists of a concept–relation pair; the input 'canary CAN' is represented by darkening the active input units. The network is trained to turn on all those output units that represent correct completions of the input pattern. In this case, the correct units to activate are 'grow', 'move', 'fly' and 'sing'.

Connectionist model

A form of computational model used to understand cognitive processes by simulating the flow of activation among simple, neuron-like processing units through weighted, synapse-like connections.

Backward Propagation

Backpropagation is a method used in artificial neural networks to calculate a gradient that is needed in the calculation of the weights to be used in the network. Backpropagation is shorthand for "the backward propagation of errors," since an error is computed at the output and distributed backwards throughout the network's layers. It is commonly used to train deep neural networks.

<https://en.wikipedia.org/wiki/Backpropagation>

Connectionist models aim to provide an explanation for how concepts and categories are acquired in a graded fashion from experience.

Acquiring Semantic Representations

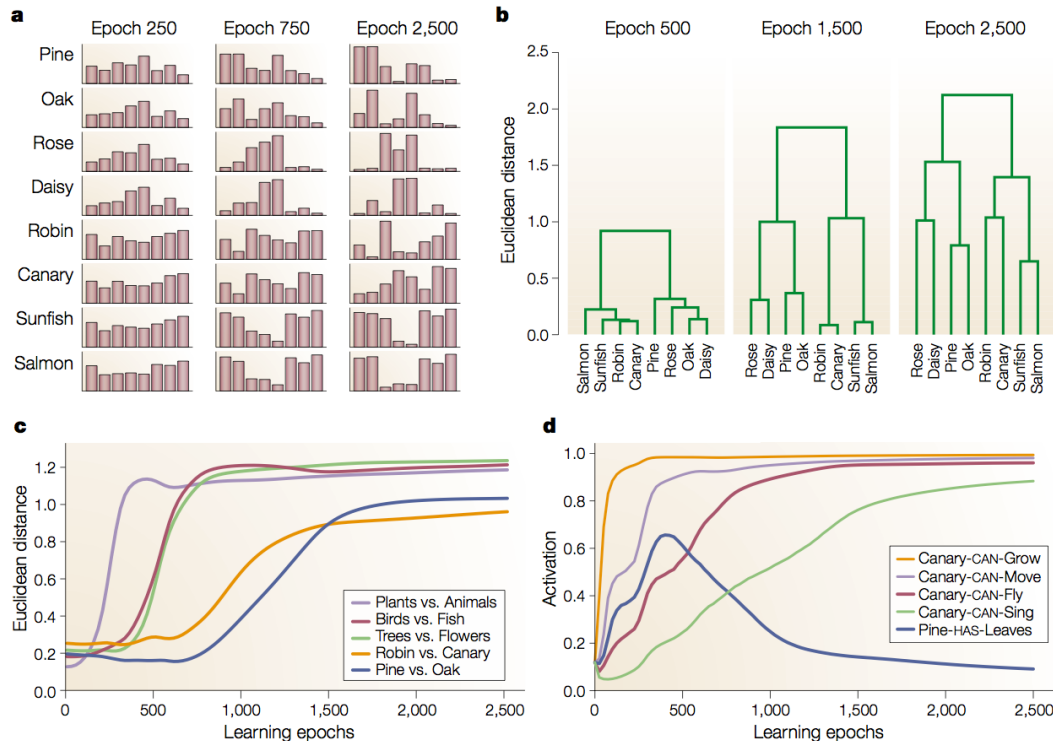
Distributed Networks

a | Patterns of activation in the feedforward network representing the eight objects (e.g., pine, oak, salmon) at three points in the learning process (epochs 250, 750, 2,500).

b | A hierarchical clustering analysis was used to visualize the similarity structure in the patterns of activation. Early in learning, the patterns are relatively undifferentiated; the first difference to appear is between plants and animals. Later, the individual concepts are differentiated, but a hierarchical organization remains showing a clear differentiation at both the superordinate (plant–animal) and intermediate (bird–fish/tree–flower) levels.

c | Pairwise distances between representations of groups of concepts or individual concepts, illustrating the continuous but stage-like character of progressive differentiation.

d | The network's performance in activating various properties of some objects indicating that correct performance is acquired in a general-to-specific manner, and tracks the differentiation of concepts shown in c. Note the activation of 'leaves' when the network is probed with 'pine-HAS'. This shows an inverted 'U'-shaped developmental course, capturing the 'illusory correlations' or incorrect attributions of typical properties.



The toy example suggests that learning of concepts can be acquired over time through learning of features. Crucially, it leads to interesting developmental patterns (global-to-basic) and errors (over-generalization). One should, however, note that learning by supervised learning with explicit, external feedback as in this example is not very plausible...

Acquiring Semantic Representations

Modern language models (including large language models, LLMs) use self-supervised learning to predict the next token in a sentence, given the surrounding context.

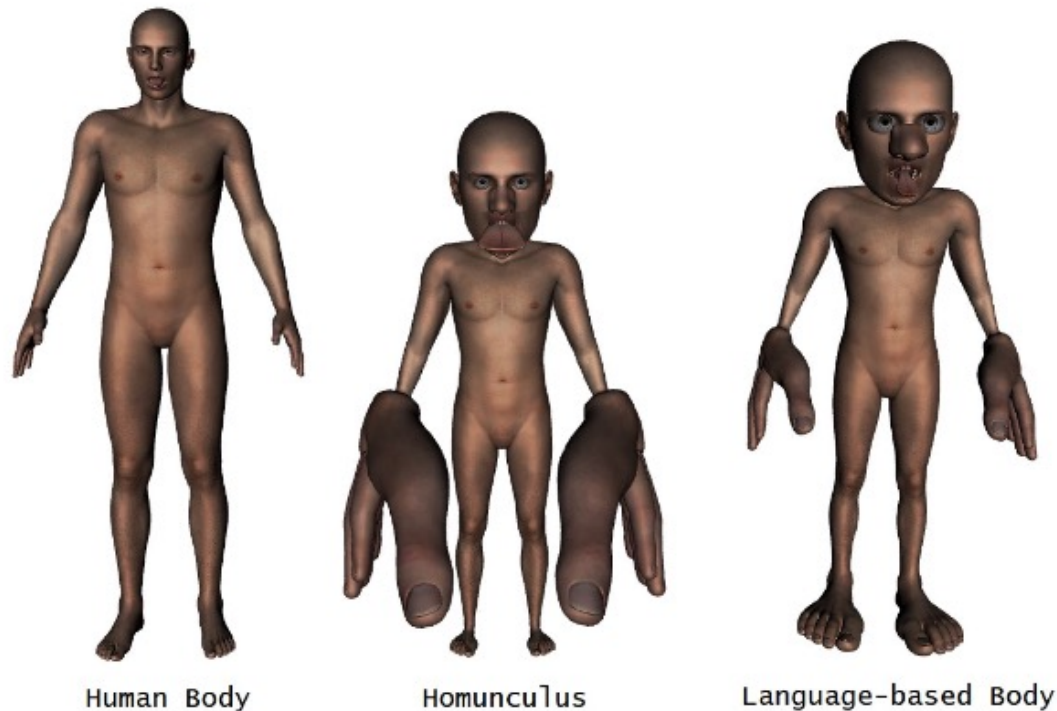
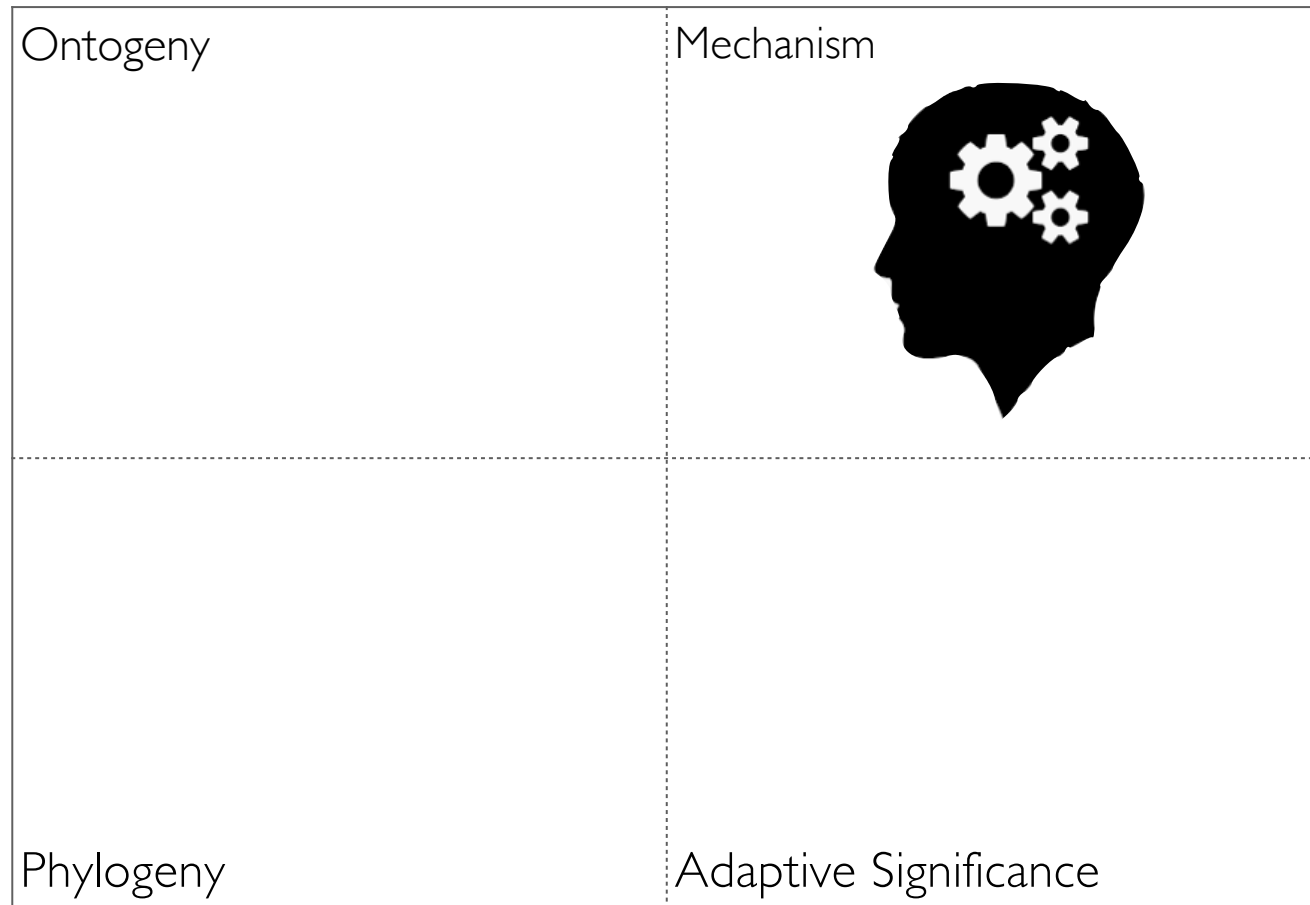


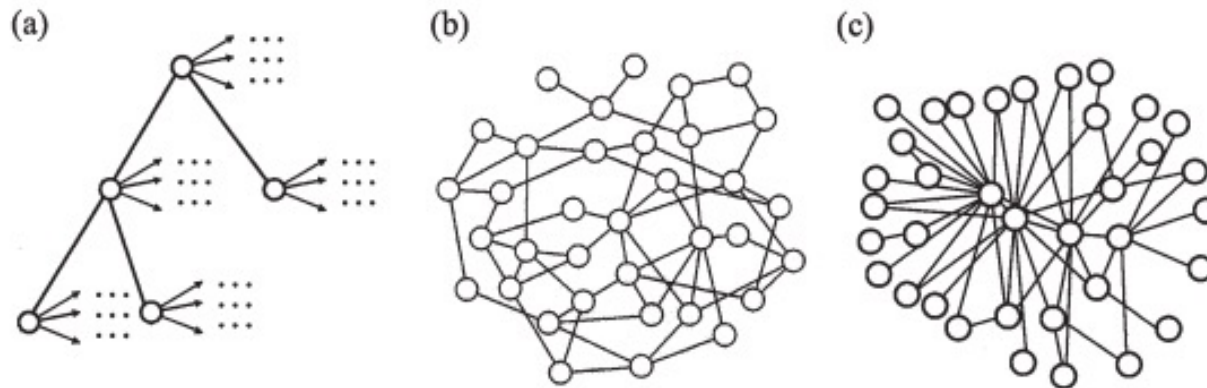
Figure 4. *Left* Actual human body proportions. *Middle* Sensory representational size proportions (i.e., the sensory homunculus (1)). *Right* Word frequency proportions. These figures were created by computing the relative (distorted) surface area of each body part for stimulations (Homunculus) and word frequencies (Language-based body), and mapping them on the “Genesis 2 Male” model in Daz 3D (<https://www.daz3d.com/>). Yet, because the morphing software used is based on volume rather than on surface area, we further adjusted our computation to the different ratios of surface area to the volume of distinct body parts⁴².

“(...) it has been shown that we can gauge city sizes by analyzing their respective word frequencies in corpora. (...) the convergence between language and physical properties of the stimuli clearly breaks down for the human body (i.e., more relevant and functional body parts are not necessarily larger in size). (...) This demonstrates that the surface-level statistical structure of language opens a window into how humans represent the world they live in, rather than into the world itself.”

Semantics



Cognitive Models of Semantic Representations



Different proposals for network models of semantic representations:

(a) tree-structured hierarchy: Collins and Quillian (1969) proposed that people have and search efficiently inheritance hierarchies to retrieve or verify facts such as “Robins have wings” and showed that reaction times of human subjects matched qualitative predictions of this model (bird vs. robin).

(b) arbitrary, unstructured graph: whereas Collins & Loftus (1975) propose connections are based on personal experience (not logic), and this could better account for effects of specific items (robin vs. ostrich). Associations as underlying mechanism of spreading activation and priming

(c), a scale-free, small-world graph: semantic networks estimated from large linguistic corpora have a small-world structure (most nodes are not neighbours of one another, but can be reached from every other by a small number of steps) and such patterns are compatible with a process of preferential attachment (more highly connected nodes are more likely to acquire new connections; Steyvers & Tenenbaum, 2005)

Steyvers, M., & Tenenbaum, J.B. (2005). Graph theoretic analyses of semantic networks: Small worlds in semantic networks. *Cognitive Science*, 29, 41-78

Spreading Activation

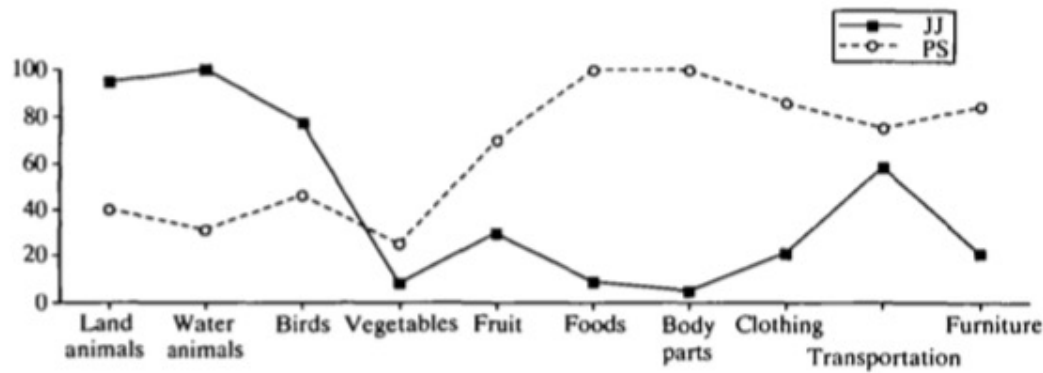
A method for searching associative networks, neural networks, or semantic networks. The search process is initiated by labeling a set of source nodes (e.g. concepts in a semantic network) with weights or "activation" and then iteratively propagating or "spreading" that activation out to other nodes linked to the source nodes. Most often these "weights" are values that decay as activation propagates through the network.

Semantic Priming

Priming is an implicit memory effect in which exposure to one stimulus influences a response to another stimulus. The seminal experiments of Meyer and Schvaneveldt in the early 1970's. Their original work showed that people were faster in deciding that a string of letters is a word when the word followed an associatively or semantically related word. For example, NURSE is recognized more quickly following DOCTOR than following BREAD.

The Neural Basis of Semantic Representations

Category-specific deficits (double dissociations from lesion studies)



Caramazza and colleagues have suggested that evolutionary pressures resulted in specialised (and functionally dissociable) neural circuits dedicated to processing, perceptually and conceptually, different categories of objects (i.e., Domain-Specific hypothesis). The hypothesis suggests specific categories for which rapid and efficient identification could have had survival and reproductive advantages: including 'animals', 'fruit/vegetables', 'conspecifics', and possibly 'tools'.

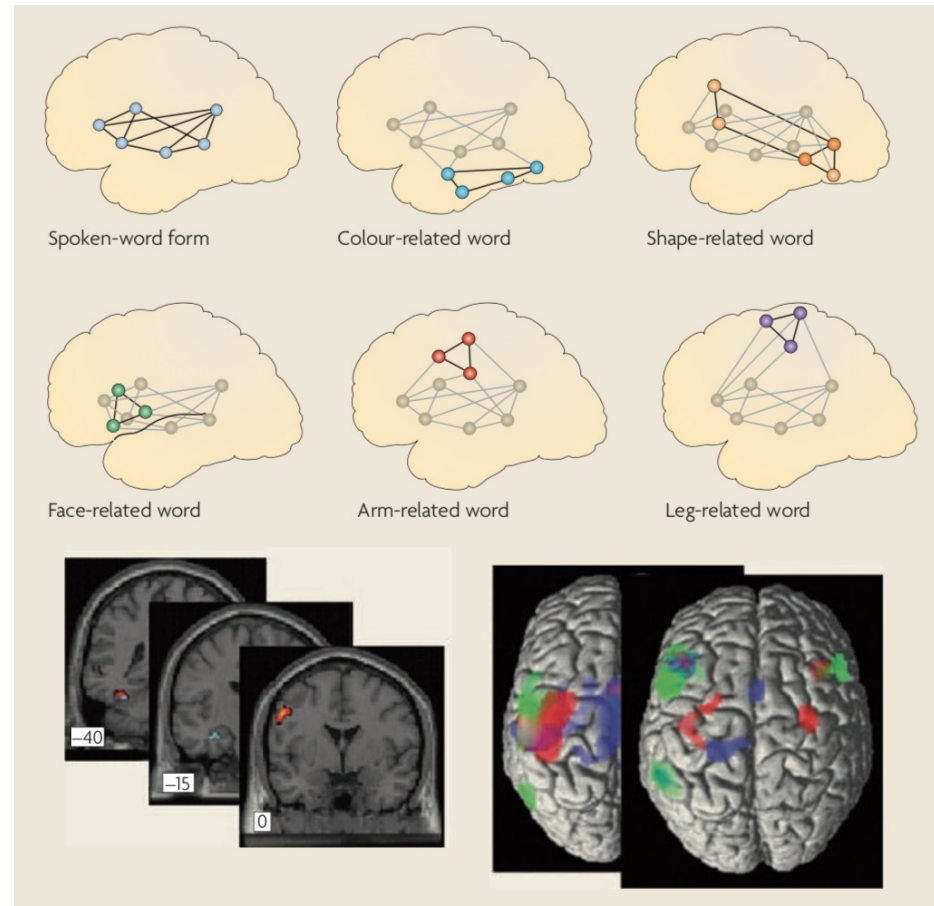
Lesion studies suggest that there are category-specific semantic deficits. More recent models (discussed in the next slides) do not dispute the categorical deficits but suggest these may be related to functional/modality characteristics associated with the categories (e.g., tools -> function and use; animals -> sensory characteristics) rather than category dedicated neural areas.

Hillis, A. E., & Caramazza, A. (1991). Category-specific naming and comprehension impairment: A double dissociation. *Brain*, 114(5), 2081–2094. <http://doi.org/10.1093/brain/114.5.2081>

The Neural Basis of Semantic Representations

Category-specific neural activation (neuroimaging)

The figure shows a model of action–perception circuits for spoken words and their meaning. **a)** Word-related circuits are located in the perisylvian language cortex, especially inferior frontal and superior temporal areas, and are strongly lateralized to the dominant left hemisphere. The learned, arbitrary links between the form of words and their meanings are provided by the coupling between these word-related circuits and semantic action–perception circuits (illustrated by different colours in the other brain diagrams). The higher-order assemblies (including both word form- and meaning-related circuits) are specific to the semantic category and store information about the actions and objects that the words are typically used to describe; **b)** Results of event-related functional MRI studies that support this model of semantic circuits.

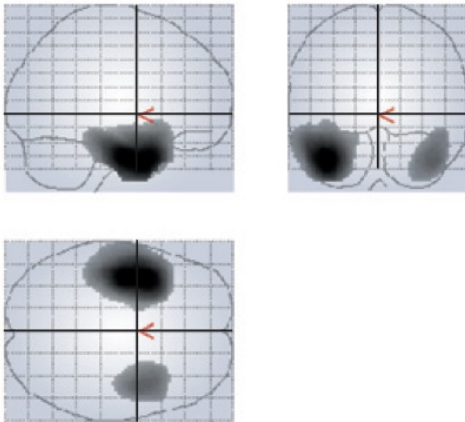


Neuroimaging studies also suggest category-specific neural activation and bolster the idea of a mapping between types of representations and specific neural circuits.

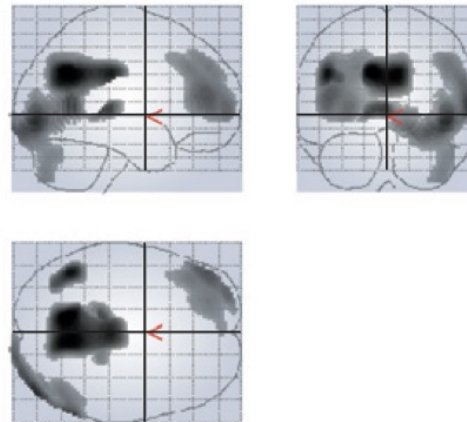
The Neural Basis of Semantic Representations

General (amodal) deficits associated with Semantic Dementia (SD)

Semantic dementia



Alzheimer's disease



Semantic Dementia

A degenerative neuropathological condition that results in the progressive loss of semantic knowledge as revealed through naming, description and non-verbal tests of semantic knowledge, resulting from disease of the anterior and lateral aspects of the temporal lobes.

There are significant differences between semantic dementia and Alzheimer's disease in measures of brain function and semantic memory. The brain areas of reduced metabolism (shown as graded grey areas in the figure above), are widespread in patients with Alzheimer's disease (AD) and include some regions that are implicated in the cortical semantic network. In the AD cases shown, however, there was little evidence of any abnormality in anterior temporal regions, which show substantial and focal hypometabolism in patients with semantic dementia (SD).

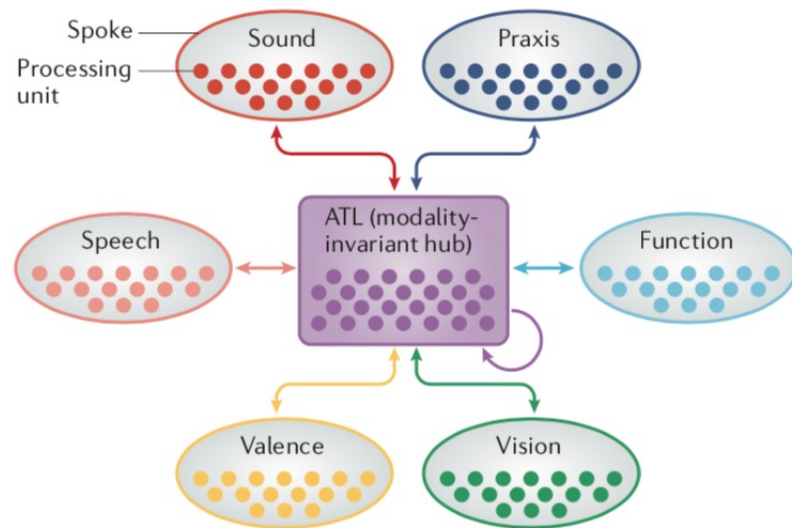
The performance of Semantic Dementia patients is significantly more impaired than AD patients on many semantic tasks (e.g., naming, verbal fluency) despite having more localised lesions.

Semantic dementia suggests that there are general (amodal) semantic deficits associated with anterior temporal function.

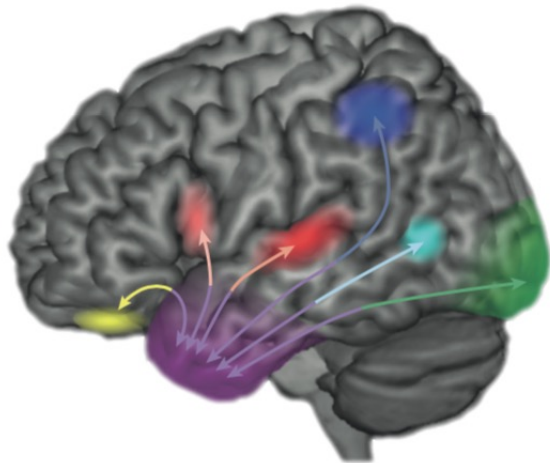
Patterson, K., Nestor, P. & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, **8**, 976-988.

The Neural Basis of Semantic Representations: Hub-and-spokes model

a Computational framework



b Neuroanatomical sketch



a | Modality-specific sources of information (spokes) are coded across a set of processing units within separate processing layers in the model. Each 'spoke' layer is reciprocally connected to a single transmodal 'hub'. The model is trained to take each of the spokes, in turn, as input and, through the hub, to reproduce the correct information across the other spokes. For example, the model is provided with the visual form of each item as input and is trained to reproduce the sounds, names, valence and other types of information that are associated with each item. The emergent result of this training is that the model forms generalizable semantic representations. The progressive, multimodal semantic impairment of patients with semantic dementia can be mimicked by gradually removing the hub connections.

b | A neuroanatomical sketch of the location of the hub and spokes is presented. The hub is located within the anterior temporal lobe (ATL) region, whereas the modality-specific spokes are distributed across different neocortical regions (the same colour coding is used as for the computational model). Each spoke communicates bidirectionally with the ATL hub through short- and long-range white-matter connections (arrows).

Summary

- **Adaptive Significance:** Knowledge (e.g., facts, causal relations) is a hallmark of intellectual performance; the cognitive system is structured such that it allows the pursuit of different (potentially incompatible) goals, suggesting the representation of abstract knowledge may be dissociable from other types of knowledge (procedural, episodic).
- **Comparative approaches:** Comparative approaches are limited in providing a picture of language-dependent, abstract knowledge; nevertheless, animal models helped understand the role of hippocampal function (central for declarative knowledge) as well as other structures (e.g., central for procedural knowledge); overall, the evidence suggests that different systems support different types of knowledge representations.
- **Development:** Evidence for developmental patterns of general-to-specific learning of concepts; current work focuses on answering how computational/learning processes can create complex cognitive representations while accounting for such developmental patterns.
- **Cognitive and neural models:** some disconnect between cognitive and neural models; there is a predominance of network models of semantic knowledge that are largely amodal but are useful to account for spreading activation and priming results from behavioural studies; current neural models, such as the hub-and-spokes model, propose both modality-specific representations (neocortex) and amodal representations (anterior temporal lobe), as well as important role for frontal cortex in cognitive control of knowledge elicitation.