January 2023: This is an early version of my hypothesis that the structure of sentential propositions was shaped by the neural architecture for hippocampal memory encoding, which I drafted and presented in a series of talks in 2022. That process, and my further reading in the history of natural language semantics and pragmatics, made me recognize that the terminology and organization of this manuscript would confuse many readers about my intended argument. My views of neural memory storage in general, and of the relationship between subjects and material individuals, have also changed slightly. Both lead me to the decision that I need to do a complete rewrite before publishing this work. However, this 'raw' version of the manuscript is dear to my heart and it seems to me that despite the confusions and errors there may still be bits which would be useful for others thinking about these problems, so I make it available here. This work was inspired by the thinking of John Macnamara, Yuki Kuroda, and my father John Lau, but especially and particularly the work of Leslie McPherson, who mercifully led me to understand what predication is. For much help (but not necessarily endorsement) I am very grateful to Tonia Bleam, Mina Hirzel, Nina Kazanina, Dave Kush, Mike McCourt, Paul Pietroski, Sandeep Prasada, Philip Resnik, Alexander Williams, and Xinchi Yu.

HIPPOCAMPAL MEMORY, INDIVIDUALS, AND SUBJECT – PREDICATE STRUCTURE

Ellen Lau
Department of Linguistics, University of Maryland

ABSTRACT

Hippocampal amnesia patients preserve knowledge from before their brain damage, and can understand and reason about new scenes and sentences, but they cannot carry forward new content beyond the current context without extensive repetition. This suggests that humans use a distinctive representational format to lastingly encode knowledge acquired from a single experience, and that the hippocampus is crucial for this kind of representation. In this paper I argue that the hippocampal format is the source of the apparent 'subject – predicate' structure of sentences. A subject-anchored data structure in the hippocampus originally developed for the purpose of navigation, providing a means for representing knowledge about *individual* locations (as opposed to general or categorical knowledge). Over time its use was extended to other kinds of individuals. Human languages were shaped to interface with this hippocampal subject-centered format because the hippocampus allows 'one-shot' encoding, such that hearers can acquire the knowledge conveyed by sentences without extensive repetition. Although the linguistic circuit itself is not the source of subject-predicate structure, grammars of natural languages provide systematic cues for how listeners should translate their meanings into that format.

INTRODUCTION

Every act of predicative judgment is a step in which a permanent store of knowledge is produced.
--Husserl (1948/1973, p. 62)

For millennia, those who carefully observed human language and thought have remarked a common intuition about sentence meanings: their underlying 'subject – predicate' structure[1]. That is, we can usually identify something that the sentence is 'about' (the subject) and the knowledge that is communicated about that thing (the predicate). In a sentence like 'The octopus ate the shark', the sentence intuitively tells us something about the octopus: that it ate the shark. Aristotle thus characterized predication as an act that unconcealed a truth about the subject, and constructed a logic that depended on this structure.

What makes these observations surprising, is that it is not immediately obvious *why* the subject – predicate format is necessary, or useful, for the purpose of expressing information about the world. A complex situation or state of affairs can be straightforwardly represented without privileging or separating out any one part of it the way the subject does in subject – predicate form. For example, if we take an objective view on the octopus-eating-shark event, the participation of the shark, and the event type itself (eating), seem just as important to our understanding of the event as the participation of the octopus. As Krifka and Musan (2012) put it:

> This does not follow from a general definition of information. For example, relational databases or sets of possible worlds, both models of information, do not presuppose any relation of aboutness.

Since it is hard to see why classical subject-predicate structure is functionally necessary, and since it is only indirectly reflected in the grammatical structure of sentences, over the years its existence has often been disputed, or its importance discounted as 'mere' pragmatics. Whether sentence pairs like "The octopus ate the shark" and "The shark was eaten by the octopus" have the same 'meaning' has been passionately debated. Frege famously wrote 'A distinction of subject and predicate finds no place in my way of representing a judgment', and many subsequent theorists of natural language semantics followed suit.

In this paper I would like to defend the classical view of Aristotle and others, that subject-predicate structure is central to an understanding of human language and thought. The fact that this format is not necessary for representing knowledge about the world—that computer databases can and do represent knowledge without subject-predicate structure—is, I think, actually the best kind of clue

[1] Many modern linguistic formalisms use the word 'subject' to refer to a grammatical property rather than one relating to meaning; thus the 'grammatical subject' does not always map to the subject in the aboutness sense, as in sentences like 'It is easy to find the post office'. This conflation of the subject with the grammatical features that (imperfectly) signal it, dates back to the beginning and is natural since they are often correlated, but it is unhelpful when we want to distinguish language from non-linguistic knowledge. Here I will use 'subject – predicate' to refer to the aboutness relation, not properties of grammatical form. Although other partial synonyms for the terms of the 'aboutness' relation exist (e.g. 'topic' – 'comment', 'theme' – 'rheme'), they have some undesirable connotations. Since the subject – predicate terminology was introduced originally as part of an investigation of perception and knowledge, it seems appropriate to continue that usage (Kuroda, 1972 gives a beautiful exposition of the terminological issues). Alternative terminology (e.g. 'external argument') can be used for grammatical correlates of subjecthood.

that its cause lies in idiosyncratic properties of humans, and not in the logic of the computational problem itself. To find the source of the subject – predicate relation, we must look to another unexpected property of human and animal memory that is not mirrored in the machines that we design, and which Aristotle could not have known or guessed: the unique role of the hippocampus in preserving knowledge.

HIPPOCAMPAL AMNESIA

Henry Molaison, known in the scientific literature by the initials 'H.M.', represents the most important single case study in modern psychology. As a young adult, Henry suffered from debilitating epilepsy--so debilitating that in 1953 he and his family decided to try an experimental surgery which would remove the hippocampal structure, a common epileptic focus, from both sides of his brain. After the surgery, Henry's seizures were indeed greatly reduced. But it soon became clear that the removal of the hippocampi had left him with severe and permanent memory deficits of a striking nature, first carefully investigated and documented by young psychologist Brenda Milner.

Milner's work with H.M. revealed a memory disorder that was profound but simultaneously extremely selective, constituting one of the strongest and most enduring arguments for the idea that human memory is non-unitary. Henry had relatively preserved memory for knowledge acquired before his surgery. He remembered his family, facts about them, and past occasions they had spent together. He had normal conceptual knowledge (e.g. what it is to be a dog, how cars function, etc.), he remembered facts about the world he had learned in school, and had preserved knowledge of language. Henry also had no problem encoding new information in the short-term: he could hold a sequence of numbers in mind for as long as the average person, and he could carry on a fairly normal conversations that required holding in mind the topic under discussion and recent comments or questions from his companion.

What Henry could not do, was to carry most forms of new information forward beyond the current context. Five minutes after a conversation was over, if his attention had shifted, he would have no recollection that it occurred. He could not remember names, or the meanings of new words, or new facts that he had been told, or new things that had happened to him. Even here, however, Milner discovered that Henry's inability to encode new information was selective. When Henry was given a challenging motor task—mirror drawing—to practice each day, his performance improved just like a normal participant's, even though each day he had no recollection that he had ever done the task before. He similarly showed normal implicit priming effects of various kinds.

Soon, other hippocampal surgery patients with the same profile of deficits were identified, confirming that H.M. was not an outlier case (Penfield & Milner, 1958). Although such surgeries obviously ceased once the functional importance of the hippocampus was realized, since this time patients with naturally-occurring bilateral hippocampal damage with a similar profile of deficits have continued to be identified[2].

---

[2] Since the hippocampus is located deep within the medial surface of the cortex and thus protected from selective injury, is rarely subject to isolated damage from stroke, and appears to support partially redundant memory functions across the two hemispheres, naturally-occurring cases of selective bilateral hippocampal damage and the corresponding memory deficit are extremely rare, explaining why they were little remarked in the historical record prior to H.M.'s surgery.

The dissociation between the ability to learn propositional knowledge vs. perceptual/motor skills disproved the common idea within behaviorist psychology that acquiring knowledge of the world was the same kind of process as acquiring habits or motor routines. The patients showed that acquiring a certain kind of knowledge clearly depended on the hippocampal circuit[3] in a way that learning perceptual patterns and motor skills did not. Exactly what kind of knowledge, is the question we turn to next.

INTERPRETING HIPPOCAMPAL MEMORY

A distinction between short-term and long-term memory was motivated long before cases like H.M.'s on various grounds, which suggested to at least some people that they might be supported by distinct neural mechanisms or regions of the brain. But because both short-term memory and existing long-term memory remained intact in hippocampal amnesia, researchers were forced to conclude that new long-term propositional knowledge could not be encoded through interactions between the short-term and long-term storage circuits alone. Surprisingly, the discovery of hippocampal amnesia seemed to require a *third* neural memory circuit: a hippocampal 'waystation' between short-term and long-term memory.

The hippocampus is known to have a different neural structure than most of the human brain: unlike most of the rest of the cerebral cortex, or 'neocortex', which under the microscope has six distinct 'layers' of varying cellular composition and input-output wiring, the hippocampus only has three or four layers. It has a distinctive and intricate recurrent circuitry, and through the neighboring entorhinal cortex has reciprocal connections to many different parts of the brain. And so these striking differences from neocortex were at least consistent with the idea that the hippocampus plays a different—if surprisingly important—role in memory.

In early years following H.M.'s case, researchers primarily focused on verifying and describing the 'waystation' role of the hippocampus, rather than explaining why the brain implemented long-term memory encoding this way. In standard hippocampal consolidation theory, features of the current situation are represented initially by distributed areas all over the neocortex (e.g., the visual properties of the context by the visual areas, the acoustic properties by auditory areas, the olfactory properties by olfactory areas...). The hippocampus creates a long-lasting 'trace' of the experience linked to all these other areas, such that any time the trace is invoked, all the linked representations from the original experience will be reinstantiated. Initially, the original representations are only indirectly linked to each other through the hippocampal trace, and if this trace is destroyed, memory of the experience is destroyed along with it. Over time (especially during sleep), by constantly reinstantiating the hippocampal trace, long-lasting direct links get formed in neocortex between the

[3] I simplify the neuroanatomy a bit in this paper for purposes of exposition. The circuit that supports what I will call 'hippocampal memory' crucially depends on adjacent entorhinal cortex as well as the hippocampus itself. Their distinct contributions are critical to a complete understanding of the neural implementation and its evolutionary history, but in this paper my goal is to characterize the properties of this memory circuit as distinct from others, and so I will simply use the shorthand 'hippocampal memory' to cover both. I follow Murry and Wise (2012) in assuming that perirhinal cortex was not historically a core component of this circuit.

‘pieces’ of the memory. This ‘transfer’ process of newly acquired knowledge from hippocampal-dependence to hippocampal-independence in the neocortex was called consolidation.

Why was such a slow and circuitous hippocampal-dependent consolidation process needed—why couldn’t the necessary links be formed immediately by neocortex itself? Broadly speaking, there were two informal intuitions. One was that it might solve a wiring problem. Humans are able to encode on-the-fly new world knowledge that combines any of our thousands and thousands of concepts and perceptual properties in different ways. To have pre-existing direct links from each ‘atom’ of the memory to every single other possible ‘atom’ would be overwhelming. With hippocampal consolidation, the pre-existing links could be initially constrained to widely-connected traces in a hippocampal ‘hub’, and with more time, the wiring of long-term memory could be reorganized such as to allow the same information to be encoded with more direct links between the parts. The other intuition was that hippocampal memory waystation buffered immediate experience because figuring out the implications of the experience for updating *generalized* knowledge required costly computation that had to be done offline. Marr (1971) wrote ‘Further classification of the information may be carried out later but, at that moment, the animal needs simply to be able to store it in its present form.’ In the same spirit, McClelland, McNaughton and O’Reilly (1995) wrote ‘the neocortex learns slowly to discover the structure in ensembles of experiences. The hippocampal system permits rapid learning of new items without disrupting this structure’. This intuition seemed to be supported by later work showing that some hippocampal amnesia patients *could* acquire new generalized knowledge when information was repeated multiple times, particularly in the case of children who suffered hippocampal damage in childhood (Vargha-Khadem et al. 1997).

The addition of a third ‘box’ in the memory model suggested the possibility that this hippocampal waystation might have its own characteristic data structure, distinct from that of working memory and long-term memory. Pursuing the intuition that the hippocampus buffered the contents of immediate experience, it was natural to suppose that hippocampal memory might be organized around episodes in time—the units of experience. In 1972, the psychologist Endel Tulving wrote an influential paper arguing for an episodic memory system for events and experiences qualitatively distinct from a ‘semantic memory’ system storing concepts and symbols. Neuroscientists quickly adopted the term ‘episodic memory’ to characterize the functional role of the hippocampus, and today this remains the most familiar and commonly used description of human hippocampus function in textbooks and materials for broad audiences. On this view, the severe memory deficits that followed from hippocampal damage could be understood as a reflection of the simple fact that all of our non-innate knowledge of the world comes to us initially through experiences/episodes.

However, a major difficulty for theories of hippocampal memory organized around episodes was recognized soon after the initial hippocampal amnesia reports. That is, that the hippocampus seems to have evolved for the function of navigating *space*, not time, and this continues to be its dominant function in most animal species. Seminal work in rats by John O’Keefe and Jonathan Dostrovsky in 1971 showed that individual hippocampal neurons selectively responded when the animal was in a particular location in space. These neurons, called place cells, have been observed in many subsequent studies with rodents, primates, and other animals; and many other studies have demonstrated spatial navigation deficits after lesions to the hippocampus or hippocampus analog in animals as varied as fish, birds, lizards, and turtles (Murray, Wise, & Graham, 2017).

These observations prompted two kinds of new hippocampal memory theories. One kind supposed that an inherently spatial data structure is extended to support non-spatial memory functions in human. For example, in a landmark 1978 book, O'Keefe and Lynn Nadel proposed the *cognitive map* theory, suggesting that hippocampal representations were fundamentally organized around maps and that the hippocampus provided the same kinds of data structures for extended uses in human non-spatial memory. The second kind of response was to hypothesize a more neutral data structure that would cut across both spatial and non-spatial functions. For example, Eichenbaum and Cohen (1995) argued that the common denominator of episodic memory and navigation was representing relations, so that the special data format of the hippocampus was 'relational memory'.

With this abbreviated history of modern theorizing on the hippocampus, we can gain some insight into why researchers in philosophy of mind and semantics have not to date recognized much import of hippocampal amnesia for their own questions about the acquisition of knowledge or the language of thought. In the early days of the discovery of patients like H.M., the literature was mainly focused on the ruling out of extreme behaviorist theories of memory, which many philosophers and semanticists may never have believed anyway. Later on, when it was gradually appreciated within neuroscience that H.M.'s case might imply the existence of a particularly important data structure for human knowledge encoding, the label that stuck was 'episodic memory', which might have sounded like a specialized cognitive function distinct from the more basic components of the acquisition of knowledge that semanticists and philosophers of mind were interested in. The subsequent connection to spatial navigation functions would not have encouraged them further. And if any of them had been paying attention when it was suggested that this region selectively supported 'relational memory', they would have found it confusing in the absence of a clear statement of what kinds of relations 'count'.

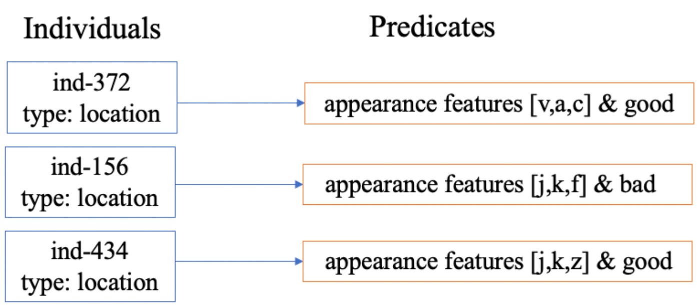
REPRESENTING INDIVIDUALS

However, there is a different way of characterizing hippocampal memory, which makes its relevance for theories of human knowledge and natural language easier to see. That is to suppose that the unique property of the hippocampal data structure is that it is centered around representations of *individuals*, as contrasted with types or general descriptions.

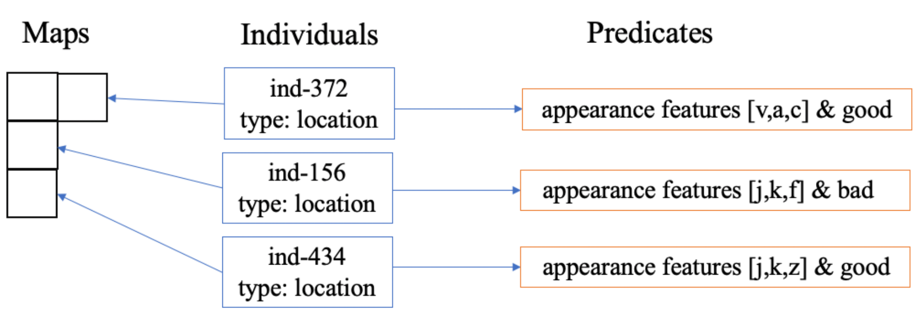
Animals can accomplish a lot with the ability to respond to, or explicitly represent, general descriptions of stuff in the world. Even very simple organisms have the capacity to, for example, initiate approach behavior upon perceiving any object of a certain range of shades of red, or to initiate flight behavior upon perceiving any approaching object of a certain range of sizes, or to avoid areas with a lot of underbrush in which predators could hide. These categorical reflexes are adaptive because they respond to some kind of generalizable properties of the animal's environment (e.g., one class of objects in the environment—what we would call lingonberries—are both red and nutritious). More complex organisms have explicit mental representations of categories, which allow them to accumulate non-perceptually based knowledge about the categories that can be used towards a more flexible range of behaviors (for example, a monkey who has a representation of the category of objects that we would call snakes might use its knowledge that they are dangerous to move away if it sees one in a tree, but could also learn a rule in an experiment that if a picture of any snake appears on the screen, it means that they should press the righthand button and they will receive a reward). All of this can be achieved without ever needing to mentally represent as *individuals*, the individual entities or locations in the world that satisfy these general descriptions[4].

On the other hand, constructing a map of locations with known properties such that one can navigate by means of those properties, depends on explicitly representing individual locations. Even if some location properties were systematically correlated (e.g. oak tree, water, good-for-finding-prey), encoding a generalized relationship between properties alone won't help the animal to use a particular location with these properties as a landmark for finding their way to something else, or to be able to find the one oak tree among many perceptual similar ones in which their nest is located. It is the relation between individual locations on the map that makes navigation possible.

It seems plausible that early amniotes (common ancestors of reptiles and mammals) first developed a circuit for representing individual locations in the precursor to the hippocampus (the medium pallium) for the purpose of map-based spatial navigation. This was perhaps the first evolved function for which explicit representation of individuals was really needed. The individual location representations would be related on the one hand to their properties—both perceptual properties that could be used to identify them, and non-perceptual properties that determined their value to the animal (e.g. 'bad / dangerous' or 'good / food source').

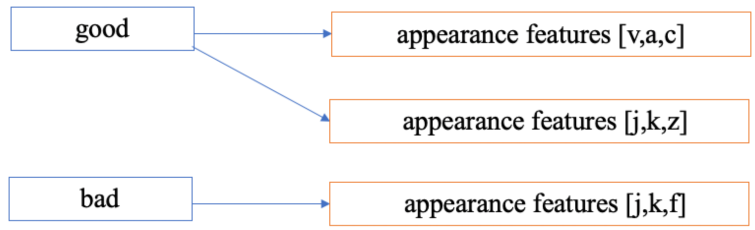


Crucially, the individual location representations had to be related on the other hand to positions in a 'cognitive map' that would afford access to spatial relationships between known locations, such that recognizing the appearance features [v,a,c] would not only tell me that I am in a good location, but can also be used to help me navigate next towards ind-434 and/or away from ind-156.

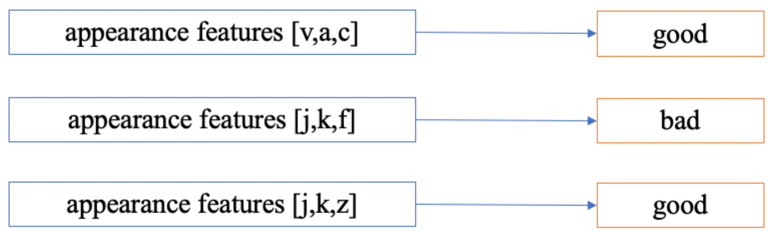


[4] If an individual in the world occasionally needs to be represented distinct from any other for some reason (e.g. an animal's mother distinct from other conspecifics), they can just be represented as a category containing a single member, on the basis of that individual's unique perceptual characteristics.

This second relationship, of the locations to the map, is what really necessitates the individual-centered format of the property information encoding. If navigation were not at issue, we could format the property information in alternative ways that could support many of the desired behaviors as long as the appearance features were always distinct enough from location to location. For example, we could center the two valence categories ‘good’ and ‘bad’, and link each of them to a list of appearance features:



Or we could center the appearance features and link each to a valence category:



If there were systematic correlations between the appearance features and valence categories, these alternative formats might be useful for making them manifest. But for map-based navigation, acquired knowledge of these properties needed to be organized around units that stand in for the individual locations, so that these units can be simultaneously related to positions in the cognitive map, the perceptual properties that allow the location to be recognized by the animal, and the properties that make the location useful or dangerous to the animal.

We can now begin to see how map-based spatial navigation in the hippocampal circuit is of interest for understanding the neural implementation of broader kinds of knowledge acquisition: the hippocampal navigation system might have been the first circuit to require systematic, explicit representation of individuals rather than categories or types. It’s likely the case that for millions of years, this circuit would be used for navigating only. But whenever a species was in an ecological niche that required learning properties on some kind of individual, this capacity could be leveraged and extended for additional purposes, such as storing knowledge about individual conspecifics for social purposes. In recent years, intracranial recordings in humans have observed hippocampal units that appear to selectively code for individual objects or people, sometimes called ‘Jennifer Aniston’ neurons (Quiroga, 2005).

INDIVIDUAL KNOWLEDGE AND RAPID ENCODING

The last crucial distinction is that encoding of knowledge about individuals can be fast and immediate, where encoding of generalized knowledge usually needs to be more gradual. Vargha-Khadem and colleagues (1997), in landmark work, showed that children who suffered hippocampal damage early in life preserved the ability to acquire generalized knowledge—doing normally in school, for example—even though they showed striking deficits in navigation and in remembering conversations, stories, and what they had done that day. Further work suggested that the repetition of information across time played a crucial role; for example, one patient showed severe recall deficits when probed on the contents of a news program after watching it once, but approached normal recall performance after watching it four times (e.g. Baddeley, Vargha-Khadem & Mishkin 2001). Similarly, Melissa Duff and colleagues have shown that adult hippocampal amnesia patients are able to acquire new 'shortcut' labels for novel objects (e.g. 'siesta man') when the objects are encountered many times in a communicative task across multiple sessions and days (Duff, Hengst, Tranel & Cohen, 2006). Thus non-hippocampal knowledge acquisition is characterized by a more gradual trajectory of consolidation across multiple experiences.

There is an obvious reason that memory circuits for acquiring generalized knowledge of types or categories should be designed to be accomplished gradually: distinct encounters with multiple individuals provide direct evidence about what properties are likely to generalize across individuals in the future, and provide a basis for inferring whether adding a new type or category is needed or not to effectively model the world (is this a single very odd horse, or a completely different species?). In contrast, when the goal is actually to acquire knowledge about a particular individual, there is little reason not to encode the observed properties as rapidly as possible—other than perceptual error, the individual's observed properties just *are* the properties of the individual at that time.

In the map-based navigation case, there might be additional functional pressures towards rapid, 'one-shot' encoding. Building a map that accurately represents the spatial relations between known locations depends on being able to accurately identify the known locations when they are re-encountered. If all the identifying properties of each location were only encoded gradually after multiple encounters, successfully constructing the map might be more challenging. It's also the case that many properties of individual locations can change rapidly across time (a food source gets consumed, a predator takes up a new lair), and therefore rapid, one-shot encoding of the location's current properties ensures that the map on which goal-directed navigation operates is up-to-date. Recent work suggests that hippocampal representations may be largely overwritten after 3-4 weeks if unused (see Barry & Maguire, 2018 for review).

The instantaneity of encoding in the hippocampal circuit is important for understanding why this circuit got extended—or 'recycled', in Dehaene and Cohen's (2007) sense—to contribute to the acquisition of knowledge beyond navigation. Although the gradual learning of the generalized knowledge system in neocortex was a feature, not a bug, once a rapid, 'one-shot' system for binding complexes of properties to individual representations was implemented for the purpose of spatial navigation, it could be used to complement and augment that system in other ways. In humans, the hippocampus has come to play a central role in acquiring many kinds of information about the world, augmenting the existing gradual-learning neocortical system. While I disagree with Marr

(1971) and McClelland et al. (1995) that this augmentation is the reason for the existence of hippocampal memory or its properties, the major insight of both works was that extending the breadth of knowledge encoded in an individualized format by the hippocampus provides organisms extra time to discover generalizations in the stored data while at rest. The evidence from children with hippocampal damage indicates that this hippocampal extension is not necessary for learning many generalizations about the world, but the evidence from adults with hippocampal damage indicate that if it's available, you come to rely heavily on it.

SUBJECT – PREDICATE

I will now suggest that certain properties of natural language sentences derive from the representational format of the hippocampal circuit for acquiring knowledge of individuals.

We can assume that the relevant properties of the hippocampal circuit predate language (and any other human-specific capacities) based on the comparative biology evidence discussed above that its structural and functional properties date back to a circuit for navigation that emerged in a common ancestor of reptiles and mammals. My guess is that the extension of the hippocampal data structure to represent knowledge of individual entities and events also pre-dated language, given observations of hippocampal and entorhinal cells that show selectivity to individual objects or conspecifics in rats and non-human primates (e.g. Deshmukh, Johnson & Knierim 2012; Sliwa, Planté, Duhamel & Wirth 2016). What this means is that when the conditions were right for language to appear, there was a pre-existing system for encoding knowledge of individuals, which—crucially—could lastingly encode new relations between existing representations instantly, after a single exposure.

My suggestion is that human language developed features that facilitated the transfer of communicated knowledge to the data structure native to the hippocampus—that is, individual 'places' to which properties are attributed. I suggest that we can understand the individual units to which knowledge is anchored in the hippocampal system as the *subject* and the potentially abstract and complex representation attributed to it as the *predicate*. This identification of subject – predicate structure with a memory system evolved for coding individuals is consistent with a longstanding philosophical tradition observing that the grammatical subject position prototypically serves to identify a particular known to or knowable by both speaker and hearer, and the predicate prototypically to indicate a general concept that exemplifies it (e.g., Strawson 1959).

On this view, the subject-predicate format of sentential propositions is native to the hippocampal circuit; not the language circuit, nor the reasoning circuit. But natural language grammars needed to produce representations that could easily be translated into that format, and this is why so many languages have (partially reliable) grammatical cues to subject. When a speaker chooses to say 'The shark was eaten by the octopus' rather than 'The octopus ate the shark', it is not just because they want to emphasize or somehow "draw attention to" the shark, it is because they want to encourage the listener to anchor the information to the hippocampal memory unit that stands in for the shark.

Since people sometimes want to communicate generalized knowledge ('Tigers are afraid of fire') as well as individual knowledge ('My son is sick'), it may seem unclear why language would preferentially interface with the hippocampal circuit rather than the generalized knowledge system in neocortex. The most obvious explanation is that the gradual learning needed for discovering the world's structure from the world itself, becomes an unnecessary hindrance when learning from the speech of trusted friends and family. If an uncle has already lived through the multiple experiences of actual tigers and fires needed to gradually acquire the general knowledge that tigers are afraid of fire, then everyone in his community can benefit from his knowledge without having to go through those experiences themselves. But if they would like to sometimes be able to acquire and retain this knowledge from a *single* speech act ('Tigers are afraid of fire') rather than needing the uncle to repeat it over and over[5], they need to use the one-shot hippocampal memory circuit[6].

This does require one innovation that wouldn't have been necessary before language: the hippocampal units originally designed to stand in for individual locations, must now be able to stand in for general kinds when this is what people want to communicate knowledge about. Although this might look a bit paradoxical, the technical implementation may not be so difficult. Certain modern theories of natural kind concepts emphasize the intertwined relationship of the kind and the individual—that part of what it is to have a concept of the 'poodle' kind is to be able to generate innumerable mental representations of individual poodles, and to know what criteria individuate them (Prasada & Hall, 2019). Prasada (2016) further observes that many people's concept of the 'dog' kind includes not only the ability to generate innumerable dog individuals according to specified individuation criteria, but also the ability to generate innumerable dog 'sub'-kinds like 'poodle' and 'terrier'. To have hippocampal units that can represent an individual dog will already require a mechanism that uses the 'dog' kind concept to support the generation of the individual reference: specifying the kind of the individual which the hippocampal unit represents. Given that the kind concept already needs to be used for individual hippocampal reference, it seems not infeasible, especially given the kind – sub-kind facts, that it could be used to support reference to the kind itself[7] for subjects of generic sentences.

Of course, just like all hippocampal knowledge, a generalized fact like 'tigers are afraid of fire' would need to be consolidated back into neocortex with existing knowledge about tigers if it is to last for the really long-term. I assume that neocortical memory does not encode knowledge in a subject – predicate format, and therefore consolidation requires some form of translation between formats. I advance no specific hypotheses about the neocortical format here, but I would speculate

[5] Note that my claim is not that repetition of utterances is not beneficial for better encoding; I am sure that it is. My claim is that nonetheless we are frequently successful in encoding the content of utterances that are not repeated.
[6] Another possible explanation is that the combinatorial representations of the generalized knowledge circuit have a more heterogeneous structure than those of the hippocampus, and thus are harder to map in a uniform manner to grammatical representations (a variant on Pietroski's 2018 proposal that lexical meanings convert polyadic predicates to monadic predicates in order to make composition easier). However, the fact that children with hippocampal damage seem able to acquire knowledge through verbal input when it is repeated, suggests that grammatical representations *can* be mapped to representations in the generalized knowledge system, but that this route is dispreferred.
[7] Interestingly, although individuals and kinds can serve as subjects for predication, there is some evidence that unspecified members of a kind cannot. In languages like Spanish, a sentence like 'Squirrels chewed through my roof' (meaning some unknown/unspecified/unquantified squirrels) is ungrammatical. Even in English where this kind of expression can be the "grammatical subject", there is some reason to think that the true subject to be predicated in memory is an individual location in space(-time), not the unknown squirrels. When 'squirrels' appears in a sentence without a spatiotemporal location that could serve as the subject, as with an individual-level predicate like 'Squirrels are annoying', 'squirrels' must now be interpreted as the kind, and not some unspecified individuals.

that its structures may be more heterogeneous than the hippocampus, having developed to represent information about various domains of the world with diverse structures.

I hope it is clear that I do not want to argue that subject – predicate format is an ideal one for encoding knowledge about the world, and I do not think that computer engineers should rush to implement their knowledge bases this way. If there were no better way to organize knowledge than subject – predicate format, then my argument would carry little force—basic functional principles would explain the subject – predicate form of sentences, not human evolutionary neurobiology. However, I doubt that this is the case. Certainly, computer engineers working on knowledge bases over the last 50 years have not spontaneously converged on this format for storage and retrieval of facts—e.g., for a fact like 'Sheila hates Peter', most systems do not anchor the fact to 'Sheila' as opposed to 'Peter'—and it is hard to see why they would. With respect to reasoning, although Aristotle ingeniously showed that the subject – predicate format *could* support many deduction patterns, 20th century logicians repeatedly established that subject – predicate format was far from necessary to support those deduction patterns, and, to the contrary was insufficient for some other important ones. Even in Kuroda's (1992) beautiful characterization of the evidence for an underlying subject – predicate format in the grammar and interpretation of Japanese sentences, he was careful to note 'There is no a priori reason why this should be the case. From the purely logical or epistemological point of view, I do not see why asserting must be made in the form conforming to the Subject – Predicate structure in the classical sense.' I think this is exactly right: there is no logical, epistemological, or grammatical reason for subject – predicate format. The reason is a biological one—but not of the form 'neurons could only do it this way', rather: 'there was a pre-existing circuit that did it this way'.

CONSEQUENCES AND QUESTIONS

The picture that has been sketched here is that the hippocampal memory data structure is a certain kind of bottleneck for encoding the knowledge communicated by sentences. It is not the *only* representational format that can interface with language—this is obvious from the fact that adult hippocampal amnesia patients like H.M. can carry on normal conversations that use syntax, lexical, and phonological information appropriately to generate structured working memory representations, and that children with early hippocampal damage seem able to use sentences to directly but gradually update long-term knowledge representations. But if listeners want to store a discrete representation of the sentence content in 'one shot', there is no option but to use the hippocampal format in which—using Aristotle's original terminology—a subject is 'accused of' a predicate (McPherson, 2002). This proposal is quite limited; it has nothing to say about the much more fundamental question of how the compositional meanings of sentences like 'The octopus ate the shark' are structured or neurally implemented. However, it has some consequences for theories of sentences and their uses, and raises some new questions.

One consequence of this view is that all sentences must specify a subject (in the underlying meaning) if they are intended to convey knowledge that should be carried beyond the current context. This contrasts with a thoughtful and influential theory (Kuroda, 1972; Ladusaw, 1994) that suggested a division between 'categoric' sentences like 'The octopus ate the shark' argued to have a subject, and 'thetic' sentences like 'There was a cat in the fireplace', argued not to have a subject[8]. However, many of the elegant insights of the categoric/thetic division can be maintained on the current view if we simply assume that the 'hippocampal subject' of a sentence like 'There was a cat in the fireplace' is not an individual of type 'cat', but an individual spatiotemporal location. This move may have seemed an unwarranted abstraction of the conception of 'subject' to these authors at the time, but it is much less so in the current framework which assumes that standing in for spatial locations is the original *purpose* of the hippocampal subjects. The kind of sentences which should not be possible on the current view are ones that do not have an understood subject at all, for example, something like 'Grows brown' with a meaning like 'things grow brown sometimes'.

This view does not predict that grammars have to include explicit or unambiguous cues to the subject, just like they do not have to contain explicit cues to other aspects of meaning; for example, not all languages explicitly signal temporal information with tense morphology. But it suggests an explanation for why many grammars do seem to include partial cues to subjecthood, e.g. particularly distinguishing an 'external' argument position in the syntax from all other 'internal' arguments. Since listeners will have to choose a subject with which to encode the sentence meaning in hippocampal memory, it is helpful, if not strictly necessary, for grammars to provide such cues for use by the producer. Understanding this restriction on memory format as the source of the grammaticization of subject may inspire different ideas about the source of other typological patterns of grammar that depend on the external/internal argument distinction.

Although in general usage the term 'topic' has a much broader scope, certain definitions of topic in theories of information structure bear a close relationship to the 'subjects' I have in mind here. Information structure would seem to be a reasonably good descriptor for the hypothesis that hippocampal memory formats information in a particular way that shapes the form of sentences. My view differs from those which assume subjects/topics are used to modulate or respond to the listener's attentional state, or to distinguish 'old' from 'new' information, but is much closer to Reinhart's (1981) classic theory of 'sentence topics'. Reinhart suggests that sentence topics determine the organization of propositions in what she calls a 'context set', which is consistent with the proposal here in which hippocampal memory is organized in terms of subjects to which individual predicates are bound. Reinhart further implies that this sentence-topic-centered structure serves the purpose of facilitating classification of information and memory retrieval. While intuitive, here I have argued that the reason for this organization is more basic; there simply is no alternative format available for one-shot memory storage. In this light, it is interesting to re-examine the question of what drives speakers' choices about which referent to cue as the subject. One plausible possibility is that, all else equal, it is in some way 'easier' for the listener to continue to add predicates to the same hippocampal subject than to shift to locate or instantiate a different one.

This view predicts that normal language comprehension should be accompanied by hippocampal activity with a certain temporal profile; at the end of each phrase denoting the predicate of the sentence, that predicate should be mapped to the hippocampal assembly that stands in for the subject. Certain kinds of so-called 'wrap-up' effects have been observed at the ends of sentences,

[8] Again, this discussion focuses on the presence of an understood subject in the meaning and not the status of 'there' as a 'grammatical subject' in the sentence form. Kuroda's examples were largely taken from Japanese, where he noted that the choice between the case-marker *wa* and *ga* could alone drive the subtle difference between a 'categoric' interpretation and a 'thetic' interpretation.

and future work could investigate more systematically whether these occur at the positions predicted by this theory. The difficulty of recovering from certain kinds of misinterpretations might also follow from whether or not encoding to hippocampal memory has occurred yet. One classic example is ambiguous sentence onsets like "While Anna dressed the baby…" that unexpectedly resolve into "While Anna dressed[,] the baby played in the crib". Readers are often able to partially recover from their surprise, correctly reanalyzing the syntax and getting the interpretation that the baby played in the crib. Strikingly, however, they erroneously continue to maintain the belief that Anna dressed the baby (Ferreira, Christianson & Hollingworth, 2001; Slattery et al., 2013). On the current view, one might relate the relative difficulty of revising interpretations to whether or not the proposition has been encoded in hippocampal memory yet. While the complex predicate is being initially constructed in working memory, it may be easy to modify, but when it appears to be complete and is encoded in the more durable hippocampal format, it may be harder to fully retract.

An interesting question is whether the subject – predicate data structure of hippocampal memory *directly* constrains production in the way I have suggested that it directly constrains comprehension. I suspect that it does not. Unlike the listener, the producer does not need to encode any new information in hippocampal memory to construct the intended message. In certain cases they may need to retrieve information from hippocampal memory to construct the message, but we can naturally understand 'retrieval' as locating information in hippocampal or neocortical memory and translating back to a more neutral working memory format. In principle, then, the pre-linguistic message they are preparing need only take on a subject – predicate structure because this distinction will usually be grammaticized in the language they are speaking, just like temporal or aspectual information need only be included in the message representation when the grammar of the language requires it (Slobin, 1996).

More broadly, one of the enduring lessons of H.M.'s case is that the same information about the world can be represented through different neural mechanisms with different formats, without our conscious awareness. If a conversation we are having is interrupted by a three-minute bathroom break, we are not conscious of switching from a working memory representation of the content before the break to a hippocampal memory representation of the content after the break. If someone references the conversation the next day, after a night of hippocampal consolidation, we are not conscious of now being able to additionally draw on a neocortical memory translation of the same content. As far as we can tell, we seem to be able to draw on many of the same deductions and inferences under any of these circumstances. These observations suggest near-equivalence in representational capacity among these different memory formats, as well as the ability to rapidly, seamlessly, and unconsciously translate any of these formats into those that are best for logical computation.

For semanticists and philosophers of mind, this multiplicity of representation offers opportunities for reconciling competing proposals about the format of the language of thought, and the sentence meanings that interface with it. Was Frege correct that the subject – predicate distinction should not be part of the representation of a judgment? I think he is wrong, if we're talking about a proposition that is represented in hippocampal memory in the hours and days after an utterance is understood. But he may be right, if we're talking about the representation of the content in working memory within the few minutes that the conversation topic is maintained, or if we're talking about the representation that eventually gets consolidated into neocortical long-term memory. Massively conjunctive theories of semantic composition with simple monadic predicates (e.g. Pietroski 2011) may correctly characterize representations that need to be formed rapidly, such as the construction of complex predicates in working memory; but Fregean saturation-style composition with polyadic predicates may be a better characterization of gradually formed long-term memory in the neocortex. File-card based 'dynamic semantics' theories that don't qualitatively distinguish the subject from other participants may be a good model of other aspects of the structure of working memory[9]. Should modern logic replace classical logic? Perhaps—if it turns out that in the mind all formats get converted to a common one for logical computation—or perhaps both can be maintained, if logical computations are sometimes carried out in a subject – predicate format that affords classical patterns of inference, and sometimes in a different format that affords additional patterns of inference. Proponents of a single language of thought have sometimes suggested that the translation required between multiple languages of thought would be too burdensome. But the patient data suggests that translation between different knowledge formats *does* occur, without our awareness. And the heft of the burden may depend very much on how different these 'languages' are from each other. For example, in the hippocampal data structure proposed here, the 'predicate' component of the structure could be almost completely isomorphic to the structure of the sentence meaning in working memory.

Conversely, the unconscious multiplicity of representation that we see manifested in hippocampal amnesia has somewhat daunting implications for interpreting empirical evidence about conceptual and linguistic-semantic representation and process—if we observe some cognitive hallmark of the structure of the sentence interpretation, which representational structure is it telling us about? There is no simple solution to this problem; nature's joints are just a bit more hidden than we knew. But psychologists have attempted for many years to develop behavioral paradigms and analysis techniques to isolate the contributions of working memory, hippocampal memory, and long-term memory across different domains, with some success (for some recent examples and discussion see e.g. Duff & Brown-Schmidt, 2017; Andermane, Joensen & Horner, 2021; Beukers, Buschman, Cohen & Norman, 2021). And with a clear neuroanatomical locus for hippocampal memory, and good candidate regions for some of the subcomponents of working memory and generalized neocortical memory (beyond the scope of this paper), I think it is not over-ambitious to hope that neural measures could be used to isolate these different representations for study in coming years.

CONCLUSION

Why do assertive sentence meanings appear to have a subject – predicate structure? Here I have suggested the answer lies not in the structure of the world itself, nor in the structure of language, but in how the human mind ended up being structured to acquire knowledge of the world, well prior to the development of language. The subject-predicate structure reflects a central aspect of human cognition, not just a minor rhetorical device, because it determines how we initially encode much of the knowledge we acquire about the world. It is not a fuzzy notion; it is rather a hard constraint on one-shot encoding of knowledge which has variable manifestations in the human languages used to convey that knowledge. And it is a unique part of our humanity because it stems from our particular evolutionary lineage—the initial marriage of individual encoding and one-shot encoding in the brains of our ancestors for the purpose of map-based navigation.

---

[9] This suggestion due to Dave Kush.