Opinion

The neural ingredients for a language of thought are available

Nina Kazanina[1,2] and David Poeppel [ID][2,3,*]

The classical notion of a 'language of thought' (LoT), advanced prominently by the philosopher Jerry Fodor, is an influential position in cognitive science whereby the mental representations underpinning thought are considered to be compositional and productive, enabling the construction of new complex thoughts from more primitive symbolic concepts. LoT theory has been challenged because a neural implementation has been deemed implausible. We disagree. Examples of critical computational ingredients needed for a neural implementation of a LoT have in fact been demonstrated, in particular in the hippocampal spatial navigation system of rodents. Here, we show that cell types found in spatial navigation (border cells, object cells, head-direction cells, etc.) provide key types of representation and computation required for the LoT, underscoring its neurobiological viability.

A hypothesis about the nature of thought

Many of us read at some point, perhaps with a mix of surprise and bewilderment, a story about a White Rabbit with pink eyes that, as it ran by a curious and stunned girl named Alice, took a watch out of its waistcoat-pocket, looked at it, and hurried on repeating 'Oh dear! Oh dear! I shall be late!' [1]. Whether or not captivated by the storyline, no reader fails to imagine the described scene, despite it being highly unpredictable and outright bizarre. The experience exemplifies that our mind can combine 'old' concepts stored in memory (linked to words) in novel ways to construct an unbounded number of thoughts, from plain and commonplace to novel and wild.

In his seminal 1975 book *The Language of Thought*, the philosopher Jerry Fodor revived, developed, and sharpened ideas originally expressed in Saint Augustine's *De Trinitate*, Thomas Aquinas' *Questiones Disputatae de Veritate*, and, later and more systematically, in William of Ockham's *Summa Logicae*: that the creativity of the mind is underpinned by a system that stores concepts symbolically and manipulates them in a structured way, using compositional rules [2–4]. Fodor named that system the LoT, underscoring similarities, although crucially not identity, of the organisation of thinking with human language. The essence of the LoT is that the mind is a computational system that operates over symbolic representations and is compositional, systematic, and productive [2,3,5]. Compositionality (and, in a related way, systematicity) refers to the principle that the meaning of a complex thought comprises the meaning of its parts and the rules that are used to combine them. The property of productivity indicates that one is able to generate novel thoughts because the system has virtually unbounded power due to its combinatorial nature over a finite set of primitives. For example, anyone who 'possesses' the concepts *John, Mary, run*, and *pinch* can entertain all of the following thoughts: *John runs*, *Mary runs*, *John pinched Mary*, and *Mary pinched John.* This outcome is expected only if thoughts are computed compositionally from primitive concepts rather than stored holistically in memory. Thoughts are expressions with a logically deducible meaning.

Highlights

The 'Language of thought' (LoT) hypothesis advanced by the philosopher Jerry Fodor posits that mental representations are compositional and productive. Thought is symbolic and new complex thoughts are constructed from more basic concepts.

The neurocognitive feasibility of the LoT has been challenged in light of the lack of demonstrable neural correlates.

We demonstrate that the key ingredients needed for a neural implementation of the LoT are indeed available, using the example of the parahippocampal spatial navigation system in rodents.

[1]University of Bristol, Bristol, UK
[2]Ernst Strüngmann Institute for Neuroscience, Frankfurt, Germany
[3]New York University, New York, NY, USA

*Correspondence:
david.poeppel@esi-frankfurt.de
(D. Poeppel).

We note that the locution 'language of thought' has led to some unfortunate misinterpretations due to the use of the word 'language'. A more neutral locution might be the 'formal system of thought': although there are notable similarities to the language system per se, the substance of the LoT does not have to be linguistic and neither do the specific formal operations. There needs to be a mapping between the LoT and linguistic cognition (something well beyond the scope of this article), but it is important to acknowledge the independent properties apparent, for example, in studying prelinguistic infants [6–8]. Note also that LoT does not exclude other potential thought formats, such as iconic representations, analogue magnitudes, and so on, anywhere in the mind/brain [9].

In current research that seeks to link the cognitive and brain sciences, the nature of thought remains a profound problem, not to say a mystery. What neurobiological approaches have to say about concepts, the construction of thoughts, and their neural implementation is still (too) distant from the insights of the psychological and cognitive sciences, including philosophy and computer science. In particular, the computational theory of mind has, in our view, not sufficiently penetrated neuroscientific methodologies to the study of thought (but see [10,11]), which we take to be a missed opportunity. The timeliness of this topic is reflected in recent publications that seek to re-examine the relevance of the LoT concept in cognition [9,10,12,13].

The LoT requires that the brain supports symbolic representations that can be combined systematically and productively, enabling the construction of novel, previously unseen combinations. However, this hypothesis was set aside as neurobiologically nonviable (for discussion, see [14]). The dominant neurobiological theories, grounded in the notion of synapses and cell assemblies as the key substrate for all knowledge [15], went in different directions. As noted by Gallistel [14] and Gallistel and Matzel [16], the widely accepted view whereby memory resides in synapses and synaptic weights presents problems: this notion of information storage renders memory a synaptic pattern, not a symbol; and such patterns are not the right natural kinds of pattern to support composition and computation (without relatively baroque adjustments and concessions to the underlying premises; but see [17–19]).

Yet, there in fact exists evidence from neuroscience for symbolic representations and computations over them [16,20]. A key piece of evidence comes from ground-breaking research on spatial navigation over the past half-century [21–23]: the organisation of the hippocampal formation and related structures in rodents and other animals demonstrates that, at the basis of spatial navigation, lies a computational system in which abstract symbolic representations enter into algebraic-like calculations.

We argue that these findings demonstrate that symbols and operations over symbols, the kind of ingredients called for by the LoT, are implemented in the brain. In examining some key findings from the spatial navigation literature, we extend Gallistel's argument that the neural machinery underpinning spatial navigation is inherently symbolic. In so doing, we demonstrate how neural cell types found in spatial navigation successfully deliver key types of representation and computation needed for the LoT, thereby bridging the substantial gap between cognitive and neurobiological views on knowledge representation. Showing that the correct ingredients are available does not force the conclusion that all thought is spatial; rather, it shows that the neural infrastructure for LoT computation is demonstrable.

Language of thought and Predicate Calculus

The LoT, foundational in the computational theory of mind [2,3,5] can be (partially) characterised by formal logics, including Predicate Calculus (PC; Box 1). The key concept of PC is that of a

Glossary

Border cell: neuron that fires when the animal is immediately adjacent to an environmental boundary (e.g., a wall or a ridge).

Boundary-vector cell: a vectorial counterpart of a border cell; a neuron that fires when the animal is located at a particular distance and direction from an environmental boundary.

(Global/complete) remapping: phenomenon whereby a place cell ensemble may reorganise itself from one environment to the next in an unpredictable way. A place cell that fires in some location in one environment may fire at a completely different location or not at all in a different environment. Global remapping enables a population of place cells to encode many different environments.

Grid cell: neuron that fires when the animal is located at one of multiple locations corresponding to vertices of a periodic triangular array that tiles the surface. The name 'grid cell' points to its function as providing a coordinate grid/frame upon which the animal can construct their cognitive map.

Head-direction cell: neuron that fires when the head of the animal is in a particular orientation in allocentric coordinates.

Landmark cell: neuron that fires when the animal is located in the immediate proximity of a usually larger and stable landmark object.

Landmark-vector cell: neuron that fires when the animal is located at a particular distance and direction from a landmark object.

Object cell: neuron that fires when the animal is in the vicinity of a (usually smaller-sized) object.

Object-vector cell: neuron that fires when the animal is located at a particular distance and direction from an object.

Place cell: neuron that fires when the animal is located in a narrowly defined region of space known as the 'place field'.

predicate, which takes one or more arguments of the type 'individual constant' (e.g., *Fodor*, *Hebb*, *Metropolitan Opera*) or 'variable' (*x, y*). LoT predicates represent a range of meanings corresponding (in language) to verbs (e.g., *RUN*, *KISS*, *LOVE*), nouns (e.g., *DOG, CONTAINER, THEORY*), adjectives (e.g., *RED, FRENCH*), adverbs (*QUICKLY*), prepositions (e.g., *UNDER, BETWEEN*), and so on. A one-place predicate denotes a subset of entities that have the quality of the predicate; for example, of all entities in the world, *RED(x)* selects a subset of entities that are red and *DOG(x)* selects those that are dogs.

Following [9], we highlight several properties of predicates (Figure 1) that we subsequently link to the discourse of neurobiology. Most predicates are abstract, in that their meaning is not determined by a set of physical criteria or surface patterns. For instance, *CONTAINER* is an overarching category comprising physically dissimilar entities: tiny to gigantic boxes, bowls, vases, or tubes; square, round, or pyramid shaped; used to hold food, clothes, or furniture; made of plastic, glass, or metal; with or without a lid, and so on. As a further example, take the predicate *DISTANCE*, which is inherently numerical and refers to the extent of some dimension between two points. Not only may the numerical values differ significantly (the points A and B may be <0.001 m, 15 m, or 44 000 m apart), so may the units of measurements, such as from meters to tones (e.g., the distance between E and F is a semitone) to years (e.g., the distance between the Earth and Sun is 8 light min), and so on.

Box 1. Formalisation of language of thought via Predicate Calculus

Which formalism is suited to represent the LoT? Logical formalisms used to analyse natural language semantics are good candidates. PC, also known as first-order logic or predicate logic [83,84], is an often-used approach that can underpin a large part of compositional productivity and systematicity (see [85] for discussion of limitations of PC).

The lexicon of PC comprises individual constants (*Hebb, Fodor*), individual variables (*x, y*), predicates (*SING, RED, NEUROSCIENTIST*), logical connectives (∧ 'AND', ∨ 'OR', ¬ 'not', → 'entails'), and quantifiers (∃ 'there exists', ∀ 'for all/ any'). These items can be combined to produce complex expressions using rules determining which combinations of lexical items constitute well-formed expressions in PC. For example, *NEUROSCIENTIST(Hebb)* is a well-formed PC formula in which the predicate *NEUROSCIENTIST* takes an individual constant *Hebb* as argument; it corresponds to the thought *Hebb is a neuroscientist*. Predicates can also take a variable as input, as in *SING(x)*, which is a PC formula equivalent to the LoT idea *x sings*.

PC also has semantic rules that assign a meaning to individual constants, variables, and predicates. Individual constants denote entities in the outside world (e.g., the individual constant HEBB denotes the person Hebb who happens to be a famous neuroscientist). A one-place PC predicate, such as *HUMAN* or *SING*, denotes a (sub)set of entities: of all entities in the outside world, the predicate returns a subset of entities that have the property designated by the predicate (i.e., entities that are human or that sing, respectively) (Figure I).

Semantic rules also make it possible to compositionally derive the meaning of any complex well-formed PC formula from its parts and to evaluate it as 'True' or 'False' relative to states of affairs in the world. Table I provides examples of complex PC formula, their semantic interpretation, and a corresponding LoT idea.

Table I. PC formula and corresponding LoT statement

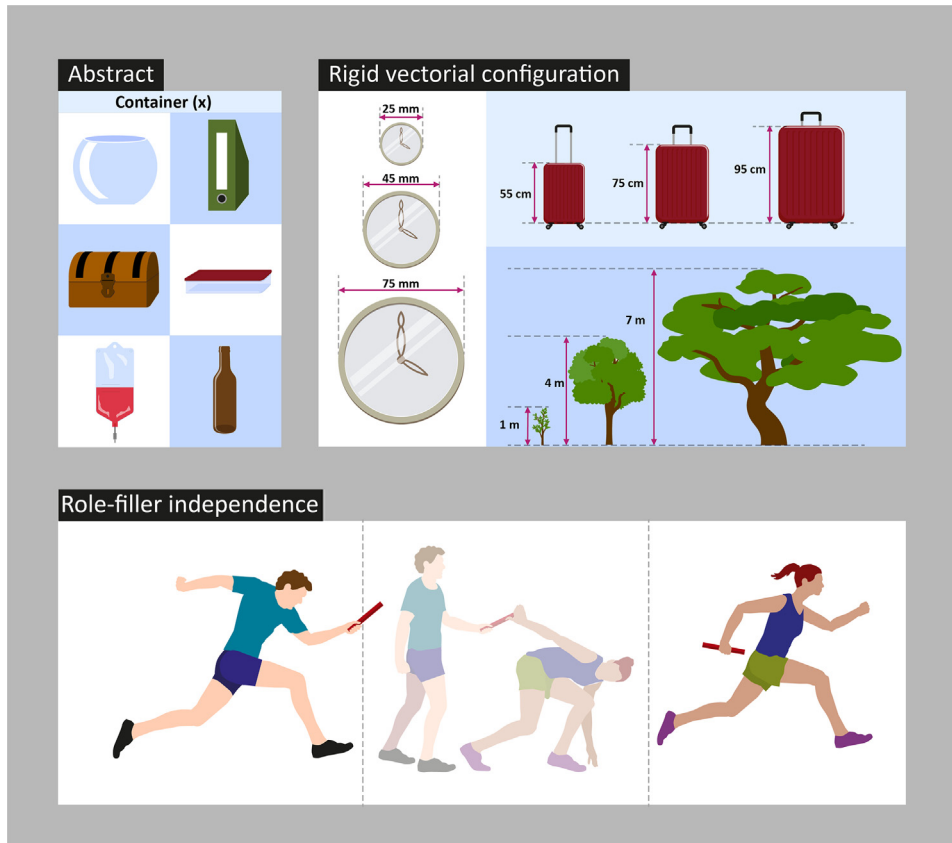| PC formula (semantic interpretation)	Corresponding LoT idea
¬Sing(Hebb) (it is not the case that Hebb sings)	*Hebb does not sing*
∃x(Neuroscientist(x) ∧ Sing(x)) (there is *x*, such that *x* is a neuroscientist and *x* sings)	*Some neuroscientists sing*
∀x¬(Neuroscientist(x) → Fish(x)) (for every *x*, if *x* is a neuroscientist then *x* is not fish)	*Neuroscientists are not fish*
¬∀x(Neuroscientist(x) → Sing(x)) (it is not the case that for every *x* that is a neuroscientist, it follows that *x* sings)	*Not every neuroscientist sings*
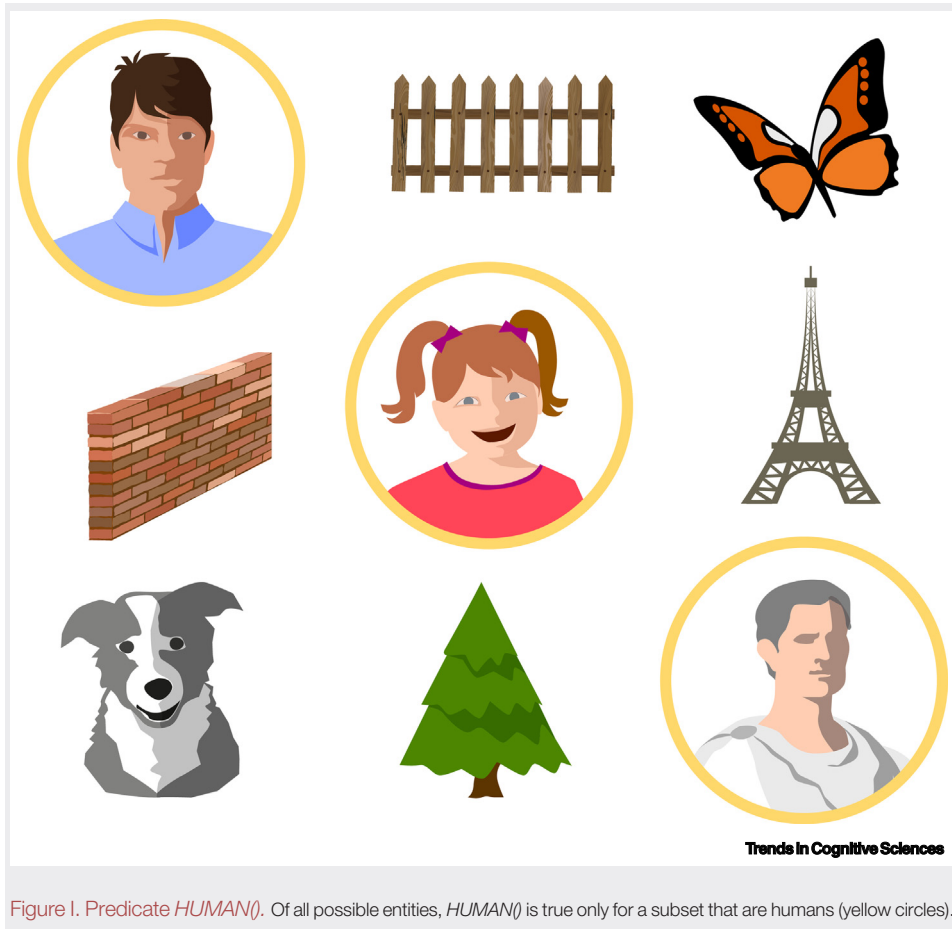
Figure I. Predicate *HUMAN()*. Of all possible entities, *HUMAN()* is true only for a subset that are humans (yellow circles).

Second, predicates has what is known as 'role-filler independence' [24,25]: predicates ('roles') are represented independently from their arguments ('fillers') and, thus, can be dynamically bound and unbound from them. For example, in a relay situation (Figure 1), during his leg, *John* can be temporarily bound to *RUN()* to obtain *RUN(John)* while Mary is bound to *PREPARE()*. When Mary's leg starts, *PREPARE(Mary)* and *RUN(John)* are unbound and *RUN(Mary)* and *WAIT(John)* are created. Thus, fillers can be bound to different roles at different times, contributing to the productivity of the LoT. Relatedly, the semantic content of the role is at least partially invariant with respect to its fillers. The core meaning of *RUN()* is maintained regardless of whether its argument is *John*, *Mary*, *rabbit*, and so on. Consequently, one can evaluate novel, previously unencountered predicate–argument combinations, such as *RUN(gleeb)*, where *gleeb* is a newly discovered creature.

Expanding the aforementioned point, because predicates maintain their core meaning as part of role-filler independence, their relative configuration in a multidimensional meaning space is stable: *DEAD(x)* remains opposite to *ALIVE(x), CHASE(x,y)* to *FLEE(x,y), ABOVE(x,y)* to *BELOW(x,y),* and *SLEEPY(x)* is more similar to *DROWSY(x)* than to *AWAKE(x)* in all contexts and for all *x*s and *y*s.

Spatial cells encode symbols and perform LOT-type computations

The linking hypothesis developed here is that the neurobiological mechanisms found in the spatial navigation system of rodents are ontologically sufficient to represent symbols and operations

Figure 1. Essential properties of language of thought (LoT) predicates. Abstractness: meaning is not determined by a set of purely physical criteria. The objects in the picture differ in shape, size, texture, and other properties, yet all qualify as 'container'. Role-filler independence: the semantic content of the predicate ('role') is (at least partially) invariant with respect to its arguments ('fillers'); for example, the predicate *RUN()* has a core meaning that holds regardless of the argument (roughly, 'move without all the feet on the ground at any given time'). Arguments can be temporarily bound or unbound to predicates; for example, *RUN()* is first bound to *John*, making *RUN(John)* true while he is running his leg, then once Mary's leg starts, *RUN (John)* becomes false and *RUN(Mary)* becomes true. Rigid vectorial configuration: *SMALL(x)* must be more similar to *MEDIUM (x)* than to LARGE*(x)* in all contexts and for *all x*s, regardless of the size of *x* or other properties.

required by the LoT. In light of the central role of predicates in the LoT, the position we advocate most critically calls for a rigorous demonstration of how such predicates (abstract, maintaining role-filler independence, and providing configurational stability) can be implemented using the neural architecture observed in spatial navigation. Of equal importance is to demonstrate that individual predicates can be combined into more complex expressions dynamically and productively.

Our linking hypothesis is stateable and viable due to trailblazing research on spatial navigation over the past half-century, pioneered by O'Keefe and Dostrovsky [21] and provided with deep theoretical foundations by O'Keefe and Nadel [26]. Various cell types have been discovered, mostly in the hippocampal formation and related structures: **place cells**, **grid cells**, **head-direction cells, border cells, landmark cells, object cells**, **landmark-vector cells**, etc. (see Glossary; Box 2). These cells enable the animal to build a cognitive map of the environment that holds information about 'places in the organism's environment, their spatial relations, and the existence of specific objects in specific places' [26]. We take the spatial navigation system to be a

Box 2. Cognitive maps and the hippocampal formation

Many animal species have impressive navigational capacities. Edward Tolman's research on navigation by rats led him to formulate cognitive map theory, whereby navigation by animals is underpinned by a map of the environment in the brain of that animal [65]. Cognitive maps are constructed during exploration and are internal representations of the external space in which distances and directions between places are encoded. They enable the animal to represent the environment comprehensively, yielding more flexible navigation than can be expected based purely on past memories (e.g., dead-reckoning) [86].

O'Keefe and Dostrovsky's 1971 discovery of place cells in the rat hippocampus that fire when the animal is in a specific location within an environment was a first step toward outlining the neural basis of cognitive maps [26]. Later, other types of cell contributing to the building of cognitive maps were found in the hippocampal formation and related structures, most notably head-direction cells, boundary cells (including border cell and boundary vector cells) and grid cells (Figure I; reviewed in [87]).

An important point is that place cells and other cells do not reflect a simple sensory activation [16,30,88]. For example, the same place cell may fire in response to a visual cue, such as a landmark, an olfactory cue, or idiothetic cues, such as when the animal moves in the dark and/or the current location is calculated using path integration. Place cell firing is independent of whether the animal is moving or stationary. This suggests that the activity of the place cell is not determined by the concurrent sensory input [26]. Rather, location is an abstract concept defined by reference to a cognitive map stored in memory. The symbolic nature of the cognitive map is emphasised by [16]: the map uses a coordinate system in which places can be identified using their coordinates (i.e., symbols for representing location).
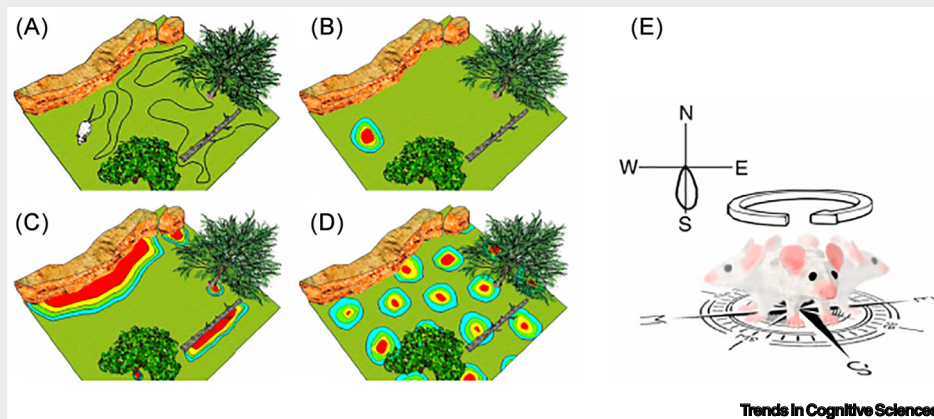

Figure I. Schematic of spatial cells in the hippocampal formation. Spatially modulated cells are recorded, for example, when a rat forages in an enclosed 2D space (A). Schematic examples of firing rate maps for a place cell (B), border cell (C), and grid cell (D). The region that yields the highest firing in the cell is indicated in red, followed by yellow, green, and so on. (E) A polar plot for a head direction cell, which fires strongly when the animal faces the preferred direction of the cell, here southward. Reproduced from [89].

compelling 'example model' for a possible LoT: despite being restricted to space, navigation has basic solutions needed for a neurobiological implementation of the LoT.

Most foundationally, the spatial navigation system features a 'lexicon' or 'inventory' of different spatial predicates encoded by single neurons: distinct cell types, such as border, object, or landmark cells, essentially implement a specific spatial predicate, *BOUNDARY(X), OBJECT(X)*, or *LANDMARK(X)*. Take border cells (also referred to as 'boundary cells') that fire when the animal is immediately adjacent to an obstacle that blocks its path [27–29]. That is, of all possible locations within the navigational space of the animal, the border cell fires only in the subset of locations where there is an environmental boundary within the reach of the whiskers of that animal (Figure 2B). This is nothing other than the implementation by the brain of the predicate *BOUNDARY(X)* in the spatial domain that returns 'True' if the receptive field X of the cell contains a boundary and 'False' otherwise. (X is a variable that uniquely encodes the position of the receptive field in 2D space, e.g., via two coordinates on the horizontal and vertical axis in the Cartesian system.)

(A)
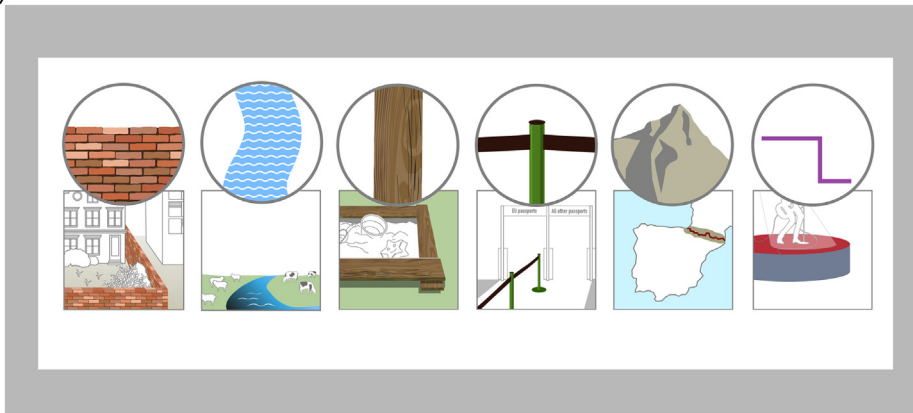
(B)
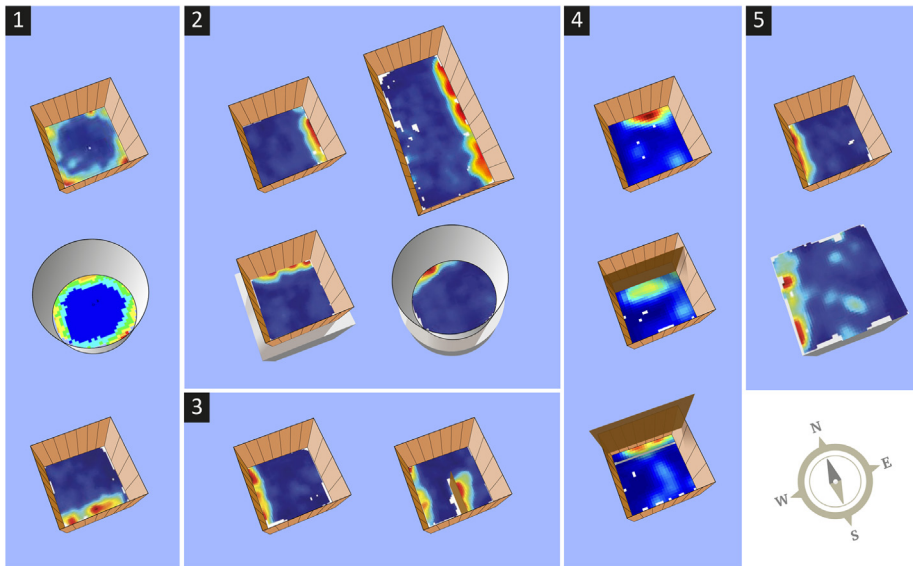
Trends in Cognitive Sciences

Figure 2. Language of thought (LoT) predicates and spatial cells. (A) The LoT predicate *BOUNDARY()* encodes an abstract meaning 'limiter, divider'. Physically dissimilar objects (circles; a wall, river, dropped level) each make *BOUNDARY()* true, as illustrated by the corresponding scenes (squares) or verbal descriptions (*The wall is the property boundary, The river is a boundary between fields*). (B) Border cells encode an abstract concept 'BOUNDARY'. (1) A border cell firing at the walls of a square enclosure (top) [29] or a circular enclosure (middle) [32]. Border cells often have a preferred direction (e.g., fire only at the south wall; bottom) [29]. (2) Border cells fire at boundaries of different geometries. The field of the cell follows the wall geometry change when a square enclosure is stretched to a rectangle (top) or changed to a circle (bottom) [29]. (3) Border cells respond to both peripheral and internal boundaries. A cell that fired at the west wall (left) also fired at a newly introduced west wall (black) inside the enclosure (right) [29]. (4) Border cells encode the meaning 'obstruction to movement'. A cell was recorded in a square enclosure (top), after a new wall was added (middle) and subsequently lifted (bottom). The cell fired at the original north wall (top), and the new north wall when it obstructed the movement of the animal (middle) but not when it did not (bottom) [38]. (5) Border cells respond to diverse boundary types. A cell fires at the west wall (top) and along the west edge after the walls have been removed to expose an open surface with a 60-cm drop along the perimeter (bottom) [29].

Importantly, the predicate *BOUNDARY(X)* implemented by border cells is abstract (Figure 2B). Border cells (and the closely related **boundary-vector cells**) fire at barriers of different textures and colours; at protruding barriers or drops in the surface level; at barriers that form the periphery

of the environment of the animal or are internal to it [29–31]. That firing by boundary-vector cells reflects abstraction is underscored even in the first report [30] of such cells: '[A] boundary is an abstract concept that may reflect sensory properties of environment features such as the sight or feel of a wall or an extended edge, as well as impediments to movement'. The firing pattern of the border/boundary cell parallels the abstractness and open-endedness of the human concept 'boundary' that can be represented by many forms (Figure 2A). This demonstrates that abstract, open-textured concepts are represented by the brain.

Another compelling example of a predicate implemented in neural tissue comes from object cells that fire when the animal is located next to an object [32–34]. As with border cells, the representation computed by the object cell is abstract. The cell fires in the vicinity of any object, notwithstanding its shape, texture, or familiarity; it fires even in the dark, when no visual information about the object is available [32]. An object cell continues to fire when an object is replaced with a different object, but stops when the object is removed completely [32]. Therefore, these cells reflect the predicate *OBJECT(X)* that returns 'True' if the receptive field *X* of the cell contains an object.

Border cells, object cells, and their vectorial counterparts satisfy role-filler independence. These cells fire for any boundary or object within their receptive field, indicating that object identity (filler) is coded by a separate neuronal population (hypothesised to be in the perirhinal cortex [35,36]) than the predicate expressed by the cell (role). Each cell type encodes a specialised core meaning, '(location of) a boundary' for border cells or '(location of) an object' for object cells. This also enables generalisation: novel, previously unseen objects can be categorised as boundaries, objects, and so on. Indeed, border cells immediately fire along novel, previously unencountered boundaries [27,29] (Figure 2B), underscoring that, just like in the LoT, the membership in the category 'boundary' can be extended to novel instances that represent an impediment to the animal. Similarly, object cells fire immediately if a new familiar or unfamiliar object is introduced into the arena [32]. The content of the receptive field of a cell is productively evaluated against particular criteria and translates into the firing by that cell.

Border, head-direction, and **object-vector cells** maintain their relative vectorial configuration. They fire coherently across environments. For example, two border cells that fire along the same/opposite walls in one environment will also fire along the same/opposite walls in another environment [29]. This also holds across cell types: the angular difference from environment A to environment B is constant for different head direction cells, and also matches that for an object vector cell [37,38]. Thus, the relative configuration of the cells is maintained, akin to how the relationship between predicates is maintained in the LoT.

Further attributes found in spatial navigation crucial for the LoT hypothesis

Similar to how LoT predicates vary in whether they reflect a more elementary or more derived meaning, so do spatial cells. Some LoT predicates may be viewed as comprising several more primitive predicates; for example, *WOMEN(x)* can be viewed as *PERSON(x)∧FEMALE(x)*, and *GIRL(x)* as *PERSON(x)∧FEMALE(x)∧YOUNG(x)* (see [2] for counterarguments). Such decomposable predicates are ample, as evidenced by the linguistic lexicon, often taken as a proxy for the LoT lexicon. The spatial navigation system has a similar mix of primitive and complex cells. For example, alongside simple border and head-direction cells, Tang *et al.* report conjunctive cells that fire when the animal encounters a border while its head is turned in a specific direction [39], representing complex meanings, such as *BOUNDARY(X)∧HEAD-DIRECTION_NORTH(A)*, where *X* represents the location of the receptive field of the cell and *A* represents the head-direction/location of the animal. These conjunctive cells designate a specific, 'fixed' meaning

(i.e., boundaries encountered when the animal faces north). Other similar conjunctive cell types have been reported (e.g., place × head-direction and grid × head-direction cells [40,41]).

Whereas some nonelementary meanings may be encoded by fixed conjunctive cells, the LoT asserts a further, properly compositional mechanism that enables the creation of new complex meanings dynamically. In the PC, the production of complex meanings 'on the fly' is achieved by combining predicates using logical connectives and quantifiers (Box 1). For example, *PURPLE BOOK* is a conjunction of two predicates, *PURPLE(x)∧BOOK(x)*, whereas *NOT A PURPLE BOOK* calls for a further combination with negation, *¬(PURPLE(x)∧BOOK(x))*. These computations are dynamic and productive (i.e., applicable to novel combinations) and, hence, are not implementable via fixed conjunctive cells that conjoin specific, 'hardwired' inputs.

Importantly, there exists clear evidence that dynamic computation is available in neural tissue. Cacucci and colleagues [42] found, in addition to typical place and head-direction cells, another cell type that they named theta-modulated place-by-direction (TPD) cells. TPD cells conjunctively code both for the location and head direction of the animal. (Additionally, their firing is also theta modulated, with spikes concentrated at certain phases of the locally recorded theta-rhythm.) Therefore, using a PC-like notation, the TPD cell encodes a meaning such as, for example, *LOCATION(A, X)∧HEAD-DIRECTION_SOUTH(A)*, that is, it fires when the receptive field *X* of the cell coincides with the current location of the animal and the animal is facing south. Crucially, whereas the locational and head-directional values of a TPD cell remain fixed across repeated exposures to the same environment, they decouple in a new environment. For example, a TPD cell that fires when the animal is facing south and is in the northeast corner in one environment fires at the same head-directional signal value (facing south) but in a completely different location (e.g., in the centre) in a different environment. Therefore, the locational signal of TPD cells **remaps** similarly to how place cells remap [i.e., it changes from environment to environment enabling the encoding of location across many different environments (see later)]. Subsequently, the value of *LOCATION(A, X)* and the entire conjunction represented by the TPD cell must be computed dynamically in a given context. Thus, TPD cells are flexible and exemplify the dynamic computation of new complex meanings needed by the LoT. For instance, a single TPD-like cell can successfully compute a combination for a noun (e.g., *BOOK*) with different adjectives (e.g., *PURPLE BOOK*, *ORANGE BOOK*, etc.).

The process of remapping observed with place cells is a further aspect of neural architecture and function that exemplifies a feature essential for the LoT. (Global) remapping refers to the observation that the receptive field of a place cell changes across contexts: a place cell that fires in a particular location in one environment (e.g., north-east corner) may fire in a completely different location in another environment (e.g., the centre) or not at all [43–46]. As a result, a limited set of place cells can be efficiently reused to represent location across many different environments at different times [47]. (Without remapping, each place cell would denote a unique location in a specific environment, which is unrealistic if the animal needs to encode many different locations and environments.) More generally, besides encoding different contexts in memory [45,48,49], remapping may enable transient representation of the ever-changing entities relevant in the current context (i.e., entities that are active in working memory and can be inputs into further computation).

Concluding remarks

The types of ingredient for the representations and computations posited in the LoT framework can be found in the spatial navigation system. This system features a lexicon of different spatial predicates encoded by different cell types. These spatial predicates are similar to LoT predicates in that they show abstraction from physical properties, role-filler independence, and relative

configurational stability. Spatial predicates can also be dynamically bound into more complex meanings.

The arguments in this article are based on the spatial navigation literature in rodents. However, spatially modulated cell types have also been found in the hippocampus and parahippocampal regions in other species, including bats [50,51], non-human primates [52,53], and humans [54–56], indicating that the spatial navigation system is conserved in evolution. Interestingly, compared with rodents, these structures in primates contain a much larger proportion of cells that respond to objects [52,53]. Notably, the human hippocampus contains cells that fire to unique concrete entities, such as a well-known person or landmark [57,58]. These cells encode an abstract concept and respond to highly varied physical signals that activate the concept (e.g., a 'Halle Berry' cell responds to different photographs of the actress, line drawings of her face, her photographs as Catwoman, the spelled words 'Halle Berry' - but not to other women's photographs or drawings). These so-called 'concept cells' correspond, roughly, to individual constants in the LoT. Similarly, cells with firing patterns that correspond to categories such as 'conspecifics' were found in the amygdala in monkeys [59], or to 'animal' or 'rodent' were found in humans [60] and can be considered as neural counterparts for the eponymous LoT predicates. Finally, as argued in [61], mirror neurons discovered in the premotor cortex of rhesus monkeys [62,63] and since found in other brain areas and species, including humans [64], can be argued to exemplify abstract predicates. The same mirror neuron fires for a specific action across many different objects and agents (e.g., a 'grasping' neuron is active whether the monkey grasps a raisin, seed, piece of apple, or a non-food object, and even when grasping is produced by another monkey or a robotic hand). Therefore, such a neuron supports encoding the action of grasping independently of it arguments and corresponds to a two-place predicate $GRASP(x,y)$.

Our perspective is inspired by innovative research in human cognitive neuroscience that is rooted in Tolman's seminal work [65] and has brought neural mechanisms found in animal spatial navigation to the problem of general knowledge organisation and reasoning in humans [10,66–74]. Behrens and colleagues [67] highlight the need to separate structure from objects; such a factorisation forms the basis for filler-role independence. Frankland and Greene [10] provide a thorough perspective on LoT-relevant computations in the brain and argue that a frontoparietal control network is responsible for compositional operations over abstract variables in the LoT. However, few details are provided as to the neurobiological mechanisms. Relatedly, some authors propose that spatial map-like representations can be used to encode abstract relations and concepts [71,72,74]. However, these lines of argumentation focus on iconic representations and have largely set aside the strictly productive, compositional aspect of computation that is central to the LoT.

The approach we take emphasises the rich representational and computational capacity of single neurons. Typically, this capacity is considered to be a function of the position of a cell within a neural assembly and its synaptic connectivity [75] and does not exclude population coding; however, there are provocative recent proposals for how this can be achieved using within-neuron RNA-based computation [76]. While it is beyond the scope of this article to argue for one or the other position, we note that an essential dimension for evaluating each approach from the LoT perspective is whether it enables dynamic compositional computation over symbols. We also note that, while most animal (especially rodent) research concerns spatially modulated parahippocampal neurons, LoT representations and computations in humans need not be restricted to the hippocampal formation. In fact, even in rodents, similar cell types have been recorded in cortex [77], demonstrating that such representations are supported widely across the brain. In addition, the encoded abstract concepts are certainly not restricted to spatial

concepts, as demonstrated by the aforementioned examples of conspecific, concept cell, or mirror neurons. Therefore, our argument that the hippocampal formation in rodents can deliver many kinds of operation needed for LoT should not be mistaken for the argument that all human thought originates in the hippocampal formation and/or is spatial in nature. This latter point is further supported by the findings of single neurons in the (pre-)frontal and parietal cortex (or counterparts) that encode cardinality within an approximate number system in domestic chicks [78], crows [79], and monkeys [80]. In humans, similar neurons have been found in the medial temporal lobe of neurosurgical patients while they performed a calculation task [81]. These 'number neurons' are abstract in that they go beyond the superficial stimulus appearance and modality (see [82]).

We have discussed the most essential components of the LoT, namely one-place predicates and compositional dynamic binding via conjunction. A fuller argument will need to provide details on predicates with two or more arguments, negation, tense, and so on (see Outstanding questions). Outstanding questions notwithstanding, we suggest that an influential position in cognitive science can be mechanistically linked to neurobiology and that a longstanding rift between cognitive symbolic theories of reasoning (and language) and neurobiological theories of memory and computation can be productively bridged.

Declaration of interests

None declared by authors.

References

1. Carroll, L. (1865) *Alice's Adventures in Wonderland*, Macmillan & Co
2. Fodor, J.A. (1975) *The Language of Thought*, Harvard University Press
3. Fodor, J.A. (2008) *LOT 2: The Language of Thought Revisited*, Oxford University Press
4. Marcus, G.F. (2001) *The Algebraic Mind*, MIT Press
5. Fodor, J.A. and Pylyshyn, Z.W. (1988) Connectionism and cognitive architecture: a critical analysis. *Cognition* 28, 3–71
6. Cesana-Arlotti, N. *et al.* (1979) (2018) Precursors of logical reasoning in preverbal human infants. *Science* 359, 1263–1266
7. Ferry, A.L. *et al.* (2015) Prelinguistic relational concepts: investigating analogical processing in infants. *Child Dev.* 86, 1386–1405
8. Carey, S. (2009) *The Origin of Concepts*, Oxford University Press
9. Quilty-Dunn, J. *et al.* (2022) The best game in town: the reemergence of the language of thought hypothesis across the cognitive sciences. *Behav. Brain Sci.* Published online December 6, 2022. https://doi.org/10.1017/S0140525X22002849
10. Frankland, S.M. and Greene, J.D. (2020) Concepts and compositionality: in search of the brain's language of thought. *Annu. Rev. Psychol.* 71, 273–303
11. Frankland, S.M. and Greene, J.D. (2020) Two ways to build a thought: distinct forms of compositional semantic representation across brain regions. *Cereb. Cortex* 30, 3838–3855
12. Mandelbaum, E. *et al.* (2022) Problems and mysteries of the many languages of thought. *Cogn. Sci.* 46, e13225
13. Cavanagh, P. (2021) The language of vision. *Perception* 50, 195–215
14. Gallistel, C.R. (2017) The neurobiological bases for the computational theory of mind. In *On Concepts, Modules, and Language: Cognitive Science at Its Core* (de Almeida, R.G. and Gleitmanm, L.R., eds), pp. 275–296, Oxford University Press
15. Hebb, D.O. (1949) *The Organization of Behavior*, Wiley
16. Gallistel, C.R. and Matzel, L.D. (2013) The neuroscience of learning: beyond the Hebbian synapse. *Annu. Rev. Psychol.* 64, 169–200
17. Smolensky, P. (1991) Connectionism, constituency and the language of thought. In *Meaning in Mind: Fodor and his Critics* (Loewer, B.M. and Rey, G., eds), pp. 201–227, Blackwell
18. Eliasmith, C. (2013) *How to Build a Brain: A Neural Architecture for Biological Cognition*, Oxford University Press
19. Smolensky, P. (1990) Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artif. Intell.* 46, 159–216
20. Gallistel, C.R. (2017) Navigation: whence our sense of direction? *Curr. Biol.* 27, R108–R110
21. O'Keefe, J. and Dostrovsky, J. (1971) The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* 34, 171–175
22. Hafting, T. *et al.* (2005) Microstructure of a spatial map in the entorhinal cortex. *Nature* 436, 801–806
23. Taube, J.S. *et al.* (1990) Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *J. Neurosci.* 10, 420–435
24. Hummel, J.E. (2011) Getting symbols out of a neural architecture. *Connect. Sci.* 23, 109–118
25. Hummel, J.E. (2019) Localism as a first step toward symbolic representation. *Behav. Brain Sci.* 23, 480–481
26. O'Keefe, J. and Nadel, L. (1978) *The Hippocampus as a Cognitive Map*, Oxford University Press
27. Bjerknes, T.L. *et al.* (2014) Representation of geometric borders in the developing rat. *Neuron* 82, 71–78
28. Savelli, F. *et al.* (2008) Influence of boundary removal on the spatial representations of the medial entorhinal cortex. *Hippocampus* 18, 1270–1282

29. Solstad, T. *et al.* (1979) (2008) Representation of geometric borders in the entorhinal cortex. *Science* 322, 1865–1868
30. Lever, C. *et al.* (2009) Boundary vector cells in the subiculum of the hippocampal formation. *J. Neurosci.* 29, 9771–9777
31. Stewart, S. *et al.* (2014) Boundary coding in the rat subiculum. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20120514
32. Jankowski, M.M. and O'Mara, S.M. (2015) Dynamics of place, boundary and object encoding in rat anterior claustrum. *Front. Behav. Neurosci.* 9, 250
33. Tsao, A. *et al.* (2013) Traces of experience in the lateral entorhinal cortex. *Curr. Biol.* 23, 399–405
34. Deshmukh, S.S. and Knierim, J.J. (2011) Representation of non-spatial and spatial information in the lateral entorhinal cortex. *Front. Behav. Neurosci.* 5, 69
35. Bicanski, A. and Burgess, N. (2018) A neural-level model of spatial memory and imagery. *Elife* 7, e33752
36. Bicanski, A. and Burgess, N. (2020) Neuronal vector coding in spatial cognition. *Nat. Rev. Neurosci.* 21, 453–470
37. Taube, J.S. (2007) The head direction signal: origins and sensory-motor integration. *Annu. Rev. Neurosci.* 30, 181–207
38. Høydal, Ø.A. *et al.* (2019) Object-vector coding in the medial entorhinal cortex. *Nature* 568, 400–404
39. Tang, Q. *et al.* (2014) Pyramidal and stellate cell specificity of grid and border representations in layer 2 of medial entorhinal cortex. *Neuron* 84, 1191–1197
40. Latuske, P. *et al.* (2015) Interspike intervals reveal functionally distinct cell populations in the medial entorhinal cortex. *J. Neurosci.* 35, 10963–10976
41. Sargolini, F. *et al.* (1979) (2006) Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science* 312, 758–762
42. Cacucci, F. *et al.* (2004) Theta-modulated place-by-direction cells in the hippocampal formation in the rat. *J. Neurosci.* 24, 8265–8277
43. Bostock, E. *et al.* (1991) Experience-dependent modifications of hippocampal place cell firing. *Hippocampus* 1, 193–206
44. Muller, R.U. and Kubie, J.L. (1987) The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *J. Neurosci.* 7, 1951–1968
45. Kubie, J.L. *et al.* (2020) Is hippocampal remapping the physiological basis for context? *Hippocampus* 30, 851–864
46. Latuske, P. *et al.* (2018) Hippocampal remapping and its entorhinal origin. *Front. Behav. Neurosci.* 11
47. Moser, M.B. *et al.* (2015) Place cells, grid cells, and memory. *Cold Spring Harb. Perspect. Biol.* 7, a021808
48. Lisman, J. *et al.* (2017) Viewpoints: How the hippocampus contributes to memory, navigation and cognition. *Nat. Neurosci.* 20, 1434–1447
49. Colgin, L.L. *et al.* (2008) Understanding memory through hippocampal remapping. *Trends Neurosci.* 31, 469–477
50. Ulanovsky, N. and Moss, C.F. (2007) Hippocampal cellular and network activity in freely moving echolocating bats. *Nat. Neurosci.* 10, 224–233
51. Rubin, A. *et al.* (2014) Encoding of head direction by hippocampal place cells in bats. *J. Neurosci.* 34, 1067–1080
52. Rolls, E.T. *et al.* (2005) Object, space, and object-space representations in the primate hippocampus. *J. Neurophysiol.* 94, 833–844
53. Eifuku, S. *et al.* (1995) Neuronal activity in the primate hippocampal formation during a conditional association task based on the subject's location. *J. Neurosci.* 15, 4952–4969
54. Kunz, L. *et al.* (2021) A neural code for egocentric spatial maps in the human medial temporal lobe. *Neuron* 109, 2781–2796
55. Jacobs, J. *et al.* (2013) Direct recordings of grid-like neuronal activity in human spatial navigation. *Nat. Neurosci.* 16, 1188–1190
56. Ekstrom, A.D. *et al.* (2003) Cellular networks underlying human spatial navigation. *Nature* 425, 184–188
57. Quiroga, R.Q. *et al.* (2005) Invariant visual representation by single neurons in the human brain. *Nature* 435, 1102–1107
58. Quian Quiroga, R. *et al.* (2009) Explicit encoding of multimodal percepts by single neurons in the human brain. *Curr. Biol.* 19, 1308–1313
59. Gothard, K.M. *et al.* (2007) Neural responses to facial expression and face identity in the monkey amygdala. *J. Neurophysiol.* 97, 1671–1683
60. Mormann, F. *et al.* (2011) A category-specific response to animals in the right human amygdala. *Nat. Neurosci.* 14, 1247–1249
61. Gallistel, C.R. (2011) Prelinguistic thought. *Lang. Learn. Dev.* 7, 253–262
62. Rizzolatti, G. *et al.* (1996) Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.* 3, 131–141
63. Gallese, V. *et al.* (1996) Action recognition in the premotor cortex. *Brain* 119, 593–609
64. Mukamel, R. *et al.* (2010) Single-neuron responses in humans during execution and observation of actions. *Curr. Biol.* 20, 750–756
65. Tolman, E.C. (1948) Cognitive maps in rats and men. *Psychol. Rev.* 55, 189
66. Tavares, R.M. *et al.* (2015) A map for social navigation in the human brain. *Neuron* 87, 231–243
67. Behrens, T.E.J. *et al.* (2018) What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron* 100, 490–509
68. Whittington, J.C.R. *et al.* (2020) The Tolman-Eichenbaum machine: unifying space and relational memory through generalization in the hippocampal formation. *Cell* 183, 1249–1263
69. O'Keefe, J. (1996) The spatial prepositions in English, vector grammar and the Cognitive Map theory. In *Language and Space* (Bloom, P. *et al.*, eds), pp. 277–316, MIT Press
70. O'Keefe, J. *et al.* (2003) Vector grammar, places, and the functional role of the spatial prepositions in English. In *Representing Direction in Language and Space*, pp. 69–85, Oxford University Press
71. Viganò, S. and Piazza, M. (2020) Distance and direction codes underlie navigation of a novel semantic space in the human brain. *J. Neurosci.* 40, 2727–2736
72. Bellmund, J.L.S. *et al.* (2018) Navigating cognition: spatial codes for human thinking. *Science* 1979, 362
73. Bellmund, J.L.S. (2020) Piecing together cognitive maps one dimension at a time. *Neuron* 107, 996–999
74. Constantinescu, A.O. *et al.* (1979) (2016) Organizing conceptual knowledge in humans with a gridlike code. *Science* 352, 1464–1468
75. Poo, M. *et al.* (2016) What is memory? The present state of the engram. *BMC Biol.* 14, 40
76. Akhlaghpour, H. (2022) An RNA-based theory of natural universal computation. *J. Theor. Biol.* 537, 110984
77. Long, X. and Zhang, S.J. (2021) A novel somatosensory spatial navigation system outside the hippocampal formation. *Cell Res.* 31, 649–663
78. Kobylkov, D. *et al.* (2022) Number neurons in the nidopallium of young domestic chicks. *Proc. Natl Acad. Sci. USA* 119, e2201039119
79. Ditz, H.M. and Nieder, A. (2015) Neurons selective to the number of visual items in the corvid songbird endbrain. *Proc. Natl. Acad. Sci. U. S. A.* 112, 7827–7832
80. Nieder, A. and Merten, K. (2007) A labeled-line code for small and large numerosities in the monkey prefrontal cortex. *J. Neurosci.* 27, 5986–5993
81. Kutter, E.F. *et al.* (2018) Single neurons in the human brain encode numbers. *Neuron* 100, 753–761
82. Nieder, A. (2016) The neuronal code for number. *Nat. Rev. Neurosci.* 17, 366–382
83. Tarski, A. (1933) Einige Betrachtungen über die Begriffe der ω-Widerspruchsfreiheit und der ω-Vollständigkeit. *Monatshefte für Mathematik und Physik* 40, 97–112
84. Frege, G. (1879) Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought. In *From Frege to Gödel: A Source Book in Mathematical Logic* (van Heijenoort, J., ed.), pp. 1–82, Harvard University Press
85. Gamut, L.T.F. *et al.* (1991) *Logic, Language, and Meaning, Volume 1: Introduction to Logic*, University of Chicago Press
86. Buehlmann, C. *et al.* (2014) Desert ants locate food by combining high sensitivity to food odors with extensive crosswind runs. *Curr. Biol.* 24, 960–964
87. Grieves, R.M. and Jeffery, K.J. (2017) The representation of space in the brain. *Behav. Process.* 135, 113–131
88. Marozzi, E. and Jeffery, K.J. (2012) Place, space and memory cells. *Curr. Biol.* 22, R939–R942
89. Poulter, S. *et al.* (2018) The neurobiology of mammalian navigation. *Curr. Biol.* 28, R1023–R1042