Perspective article

Structure learning and the posterior parietal cortex

Christopher Summerfield*, Fabrice Luyckx, Hannah Sheahan

Department of Experimental Psychology, University of Oxford, Anna Watts Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK

ARTICLE INFO

ABSTRACT

We propose a theory of structure learning in the primate brain. We argue that the parietal cortex is critical for learning about relations among the objects and categories that populate a visual scene. We suggest that current deep learning models exhibit poor global scene understanding because they fail to perform the relational inferences that occur in the primate dorsal stream. We review studies of neural coding in primate posterior parietal cortex (PPC), drawing the conclusion that neurons in this brain area represent potentially high-dimensional inputs on a low-dimensional manifold that encodes the relative position of objects or features in physical space, and relations among entities in abstract conceptual space. We argue that this low-dimensional code supports generalisation of relational information, even in nonspatial domains. Finally, we propose that structure learning is grounded in the actions that primates take when they reach for objects or fixate them with their eyes. We sketch a model of how this might occur in neural circuits.

1. Introduction

As they learn about the world, humans and other animals form new mental representations. During human development, these representations pertain to basic items and categories (e.g. "dog", "table") but in adulthood a rich repertoire of more abstract concepts is formed (e.g. "commerce", "friendship"). This ensemble of categories and concepts lays the foundation for our knowledge of the world (Murphy, 2002). For decades, psychologists have made assertions about how representations (or mental schemas) are used by higher cognitive processes involving memory, reasoning and language (Schank and Abelson, 1977). In parallel, neurobiologists have studied the organisation of representations in the brain, by measuring the tuning properties of single neurons or charting gross patterns of category selectivity with neuroimaging methods. However, there is a notable lack of consensus about how new mental representations are acquired *de novo* from natural sensory experience, and building systems that learn rich concepts remains a key goal in artificial intelligence (AI) research (Hassabis et al., 2017; Lake et al., 2016).

In this article, we discuss how humans (and potentially other animals) might learn the abstract representations that support intelligent behaviour. We begin by reviewing an emerging theory of neural coding based on contemporary deep learning models (LeCun et al., 2015). These models excel at image classification and acquire representations that resemble those in the primate brain. However, we argue that standard deep learning architectures may struggle to learn the sorts of

rich abstractions that support human intelligence, because they do not explicitly encode *relational* information among stimuli. In subsequent sections, we review evidence that neurons in the posterior parietal cortex (especially lateral intraparietal area LIP) code for object relations in physical space and go on to suggest that these dorsal stream representations may form a basis set for learning about the structure of more abstract spaces, and potentially play a role human concept learning.

2. Representation learning in deep neural networks

Recent advances in deep learning have offered a new theory of representation learning for biological brains (Kriegeskorte, 2015; Yamins and DiCarlo, 2016). Researchers have examined the representations formed by deep neural networks that are trained with supervision to perform challenging object classification tasks and compared them to observations in the ventral stream of monkeys and humans. Promisingly, the early layers of deep networks form filters that are selective to orientation and spatial frequency, resembling the tuning of V1 cells (Yoshinski et al., 2015), and the distribution of object and category selectivity in the hidden layers resembles that from neurophysiological recordings in area IT (Yamins et al., 2014). Moreover, patterns of representational similarity among population responses to visual categories are similar in humans, monkeys and neural networks (Khaligh-Razavi and Kriegeskorte, 2014). Reassuringly, composite measures reveal that artificial network architectures that excel at image

---

* Corresponding author.
  E-mail address: christopher.summerfield@psy.ox.ac.uk (C. Summerfield).

classification are those that most closely resemble the primate brain (Yamins and DiCarlo, 2016). In part, this convergence is expected from the fact that contemporary deep networks incorporate many known computational motifs from biological vision, including layer-wise depth, convergent connectivity, local receptivity, pooling and divisive normalisation (Krizhevsky et al., 2012).

However, a different perspective holds that deep networks are at best only a partial theory of visual representation learning in biological systems (Lake et al., 2016). Aside from noting the relative paucity of labelled data available in the real world (Marcus, 2018) many researchers have emphasised the differences in performance between humans and machine learning systems on object classification tasks. Deep networks are vulnerable to being misled with "adversarial" images that (to a human observer) show obvious category members (Brendel et al., 2017), and are also prone to classifying nonsense images as objects with high confidence (Nguyen et al., 2014). A related issue is that deep networks are highly susceptible to variation in the statistics of the images in training and test data sets. For example, the addition of small amounts of pixel noise or light bandpass filtering of test images confuses the networks appreciably more than human observers (Geirhos et al., 2018). Thus, whilst after extensive training deep networks may exhibit highly accurate object classification, they seem to lack the robustness that is characteristic of biological vision.

To understand these phenomena, it may be useful to take a step back and consider the nature of the problem that visual cognition evolved to tackle. Humans and other primates experience rich and complex visual scenes that typically contain multiple objects at once. Natural scenes are thus understood not just by their *contents* (e.g. the identity of the objects that are present) but also by their *structure* (i.e. the relationships among those objects). For example, two images of a person playing with a toy car and driving a car may both contain the items "human" and "car" but mere label assignation ignores the information about relative scale (is the person bigger than the car?) and containment (is the person inside the car?) that are crucial for interpreting these scenes (Fig. 1a). Nearly a century ago, the Gestalt psychologists defined a compact set of principles for describing the structure of visual information, focussing on the relative scale, proximity, similarity, and grouping of elements within an array (Fig. 1b). Critically, these structural descriptions are agnostic to the specific object labels assigned to scene contents. For example, a configuration that obeys the Gestalt principle of symmetry is agnostic to the size, shape or colour of the symmetric items (Wagemans et al., 2012). In contemporary parlance, Gestalt perception is the outcome of structure learning about natural scenes.

Deep networks are typically trained on datasets that are curated to ensure that each label pertains to a single identifiable object class (Fig. 1d). They are thus not placed under any explicit pressure to learn about object relations in natural scenes. As a consequence, whilst deep networks may excel at category labelling, they tend to show limited generalisation on tasks that require objects to be segregated from the background, counted, or compared in terms of their relative scale or position (see below). In other words, whilst deep networks are highly effective at learning a mapping between distributions of image statistics and a class label, they form only a superficial scene understanding that lacks the fundamental human notion of "objectness". In what follows, we discuss the computational principles that may hamper current machine learning architectures from fully understanding natural scenes and consider how it might be possible to build information processing systems that learn about scene structure in a way that more closely resembles the primate brain.

For a neuropsychologist, these failures of object and scene understanding in deep networks have a rather intuitive explanation. As outlined above, the representations formed by deep networks resemble those in primate inferotemporal cortex (Yamins and DiCarlo, 2016). However, primates have also evolved a parallel visual pathway that projects dorsally to the parietal cortex (Goodale and Milner, 1992).

Lesions that affect this pathway lead to deficits of visual cognition that resemble those exhibited by deep neural networks. For example, patients with bilateral parietal damage suffer from Balint's syndrome, a debilitating disorder that impairs global scene perception whilst preserving local object recognition (Chechlacz and Humphreys, 2014). One symptom of Balint's syndrome is simultanagnosia, a visuospatial deficit which impairs decisions about groups of objects. For example, patients may perfectly identify a lone object as a cat but fail to enumerate how many cats are present in a scene or identify which is the largest (Friedman-Hill et al., 1995). Parietal damage can also lead to constructional apraxia, a disorder characterised by an inability to build, assemble or draw objects from their constituent parts (Black and Strub, 1976). Thus, patients with parietal damage behave as if they lack a representation of visual space that would otherwise allow them to individuate distinct objects within a scene (e.g. a cat and a dog), and thus permit the relational inferences between elements that are required for rich and accurate scene descriptions (e.g. the dog is chasing the cat). This pattern of intact local but impaired global scene perception resembles that displayed by current supervised deep networks. Indeed, deep networks struggle on lab-based cognitive tasks that involve processing of multiple objects, such as change detection and multi-object tracking, both of which are associated with intact parietal function (Leibo et al., 2018). Neural networks also fare poorly on tasks that involve discriminating object arrays according to their Gestalt properties (Fleuret et al., 2011) and may fail at object counting (Wu et al., 2018). We suggest that failures occur because deep networks lack a dorsal steam pathway or an equivalent module for explicitly processing the spatial relations among objects (Fig. 1c).

3. Neural codes for relational structure

Thus far we have argued that biological systems, unlike current deep networks, can learn about relational structure among objects and use it for scene understanding. We also suggested that the dorsal stream pathway may play a role in this process. But how is this achieved at the mechanistic level, and how are the proposed computations consistent with empirically observed principles of neural coding in the primate parietal cortex?

Understanding the neural and computational basis of structure learning is a challenge that has recently come to the fore in neuroscience (Behrens et al., 2018; Bellmund et al., 2018; Tervo et al., 2016), spurred by the recognition that in order for AI systems to produce intelligent behaviours, they will need to encode and generalise abstract relational knowledge (Lake et al., 2016). Deep neural networks are powerful function approximators whose connections (weights) implicitly contain information about the structure of the data to which they are exposed, such as the probability of co-occurrence of different scene elements. However, rapid and versatile transfer of knowledge between environments requires that the system learns to *explicitly* represent the structure within data, by explicitly coding the relational structure among object parts, such as the patterns of brushstrokes that constitute a handwritten character (Lake et al., 2015).

To illustrate what is meant by an "explicit" representation, consider a deep network exposed to scenes in which one object is inside another (e.g. a ball in a cup or a human in a car). Contemporary deep networks can learn to decompose these scenes into interpretable constituents, as shown by the emergence of units that code for specific objects or code for feature variation in interpretable ways (Burgess et al., 2019). However, in order to support rapid generalisation to new scenes, the network needs to form units or assemblies that code for the concept of "inside" as well as "cup" or "book", i.e. to factorise scene knowledge into structure and contents (Behrens et al., 2018). The activity of such a unit (or assembly) coding explicitly for structure would allow the network to transfer its knowledge to previously unseen images, signalling for example that a dog is *inside* a kennel or a shirt is *inside* a drawer. This intuition has a long pedigree in the cognitive sciences (Gentner, 1983).
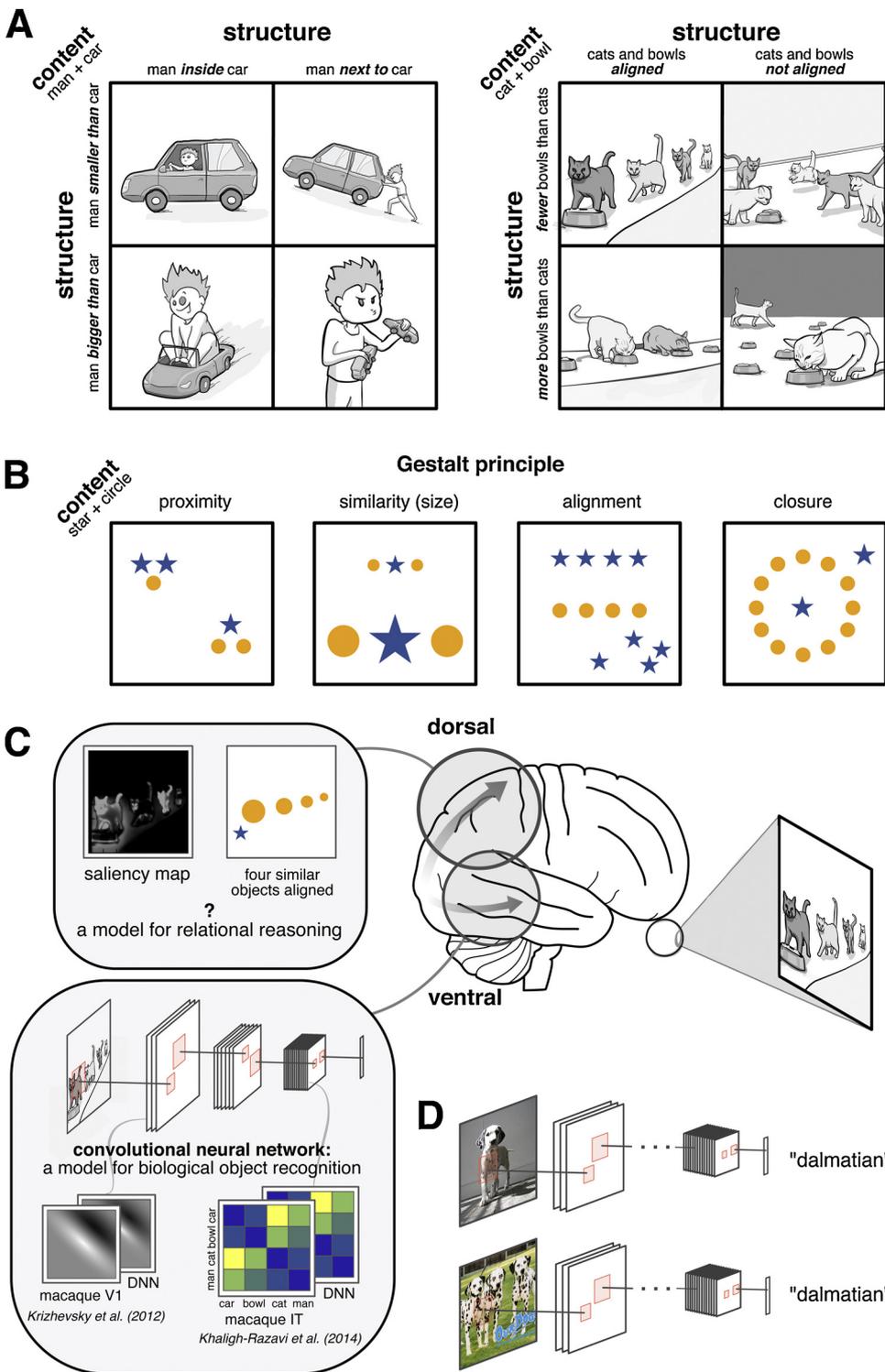
Fig. 1. The problem of structure learning. A. A visual scene is understood not just by its contents but by the relations between the objects that are present. In the example on the left, scene understanding depends on the relative size of the two objects (man > car or car > man) and containment (man in car or man outside of car). The scenes can variously be interpreted as demonstrating a daily commute, play with toys, or a vehicle breakdown. Right panel: a different example with two objects (cats and bowls) and two different structural principles: alignment and cardinality. The cats might be waiting or eating in an orderly fashion (good cats) or roaming chaotically (naughty cats); there might be more bowls than cats (cats are happy) or vice versa (cats go hungry). B. Principles of Gestalt perception offer descriptions of the relations among objects, here in terms of their relative proximity, similarity, alignment, or containment. Note that these Gestalt principles are invariant to the identity of the objects (one could replace the yellow circles and blue stars with red triangles and green squares) as well as to their gross position within the array (one could translate objects or groups within the array). C. Theory sketch. Current deep networks explain variance in ventral stream structures including V1 (they learn Gabor filters) and IT (they learn invariant object representations). However, relations among objects may be encoded in the parietal cortex (see main text for details). D. Image classification datasets, such as Imagenet (Russakovsky et al., 2015), do not distinguish relational properties of objects. For example, the top image (one dalmatian) and the bottom image (three dalmatians) both belong to the class "dalmatian". Current deep networks trained with supervision are thus not placed under any pressure to explicitly learn about object relations.

Cognitive models are often endowed with handcrafted inputs for the structure and contents of data, for example they are provided with units labelled by the researcher as corresponding to "close", "inside" or "largest" (Doumas et al., 2008). These models are geared to explain the human ability to learn about both objects and predicates, or about the syntax and semantics of natural language. However, these systems do not learn abstractions from raw input data (e.g. image pixels) as biological systems can.

One possibility is that abstract relational knowledge is grounded in inferences about physical space (Bellmund et al., 2018). One compelling version of this theory takes as its starting point the characteristic coding properties of cells in the rodent hippocampal-entorhinal system (Behrens et al., 2018). Hippocampal place cells are sensitive to single locations in allocentric space, whereas grid cells fire at multiple local positions arranged in a hexagonal lattice. In the rodent medial temporal lobe, ensembles of both cell classes tile the environment over multiple spatial scales (e.g. coarse to fine). Intriguingly, a six-fold symmetric lattice is precisely the coding pattern expected if grid cells were computing the principal components of the place cell covariance matrix produced whilst the animal navigates an open field,

under a constraint that the cells cannot produce negative firing rates (Dordek et al., 2016). The theory thus states that while place cells signal specific instances of space occupancy, the grid cells efficiently encode the way space is structured itself, i.e. the adjacency properties that are given by the geometry of the 2D (allocentric) environment. In the terms employed above, place cells code for the *contents* of experience and grid cells code the *structure* of experience. The efficiency principle (i.e. reduction of redundancy) that results from the dimensionality reduction may make these codes particularly useful for planning in a computationally frugal fashion during navigation (Stachenfeld et al., 2014).

At present, it is not clear how these principles can be implemented in a network that learns end-to-end from natural inputs. One possibility is that training a network to understand and generalise object relations will require algorithmic innovation. For example, modifying a network architecture to incorporate strong priors about potential relational patterns observed in data has allowed researchers to build systems that generalise relational knowledge within fixed domains. "Relation" networks solve visual query problems involving a set of table-top objects (Santoro et al., 2017) and with sufficient training can even solve reasoning problems, such as Raven's Progressive Matrices, that feature in human intelligence tests (Barrett et al., 2018). However, it remains unclear whether these networks will scale efficiently to more general settings. An alternative possibility is that structure learning requires a qualitatively different objective or cost function for the network (Marblestone et al., 2016). For example, one recent paper shows how distinct representations can emerge in networks trained to perform related and unrelated tasks simultaneously, arguing that this offers insight into the coding properties in distinct visual streams (Scholte et al., 2018). In sum, thus, building a system that learns abstract relational information remains a challenge for AI researchers and neuroscientists alike.

4. Neural coding in the primate posterior parietal cortex

The view outlined above is premised on an understanding of neural coding in the rodent medial temporal lobe. However, primate vision is quite different from rodent vision. Primates have forward-facing eyes and frequently direct their attention into peri-personal space, where they are adept at grasping and manipulating objects with their dextrous hands. Primates also have a highly developed saccadic system that allows them to experience natural scenes via multiple discrete samples and to combine the information over space and time, perhaps via neural integrators in the posterior parietal cortex that exhibit recurrent excitation (Gold and Shadlen, 2007). Thus, whilst rodents and primates process information in both egocentric (self-to-object) and allocentric (object-to-object) frames of reference, it seems likely that the latter is more important in the rodent and the former in the primate (Fig. 2a). This is suggested by the relatively larger size and greater functional significance of the hippocampus compared to neocortex in rodents (Fig. 2b), and the relative waking time spent engaged in locomotion (i.e. transitioning through allocentric space), which by some estimates is greater than 50% in rodents (Aragao Rda et al., 2011) but closer to 20% in macaque monkeys (Sussman and Tattersall, 1981). Instead, many primates in the wild spend much of their time foraging in arboreal environments where dextrous reaching and grasping for branches and food items (i.e. transitioning through peri-personal space) is important (Passingham and Wise, 2012). If rodents and monkeys learn about relational structure by understanding space, then the two species might vary in the extent to which their inferences are grounded in maps vs. scenes respectively. A natural question, thus, is whether neural coding in the primate parietal cortex permits inferences about both the structure and contents of a visual scene, in the same way as medial temporal lobe structures have been argued to represent the structure and contents of allocentric space in rodents.

In what follows, we argue in favour of this view. We draw upon evidence from recordings in posterior parietal cortex, a brain structure that has been extensively studied in the macaque monkey. A great deal of research has focussed on the lateral intraparietal area (LIP; Brodmann's areas 39/40), where neurons are activated by stimulation of their spatially selective response fields, and fire persistently as the monkey prepares a saccade to their preferred spatial location. LIP has been implicated in a wide range of functions (Freedman and Ibos, 2018), including spatial working memory (because activity persists when the saccadic target is removed), decision-making (because activity builds up with the probability that the saccade will be rewarded), top-down spatial attention (because locations cued as relevant elicit more activity) and in spatial priority (because activity scales with the bottom-up salience of the stimulus). Despite these diverse theories, there is good consensus that LIP neurons are spatially selective, coding for locations in egocentric space much as place cells in the rodent hippocampus code for locations in an allocentric frame of reference.

Interestingly, there is evidence that cells in an adjacent region of posterior parietal cortex code explicitly for relations among object parts, as might be predicted from the study of constructional apraxia described above. For example, Chafee and colleagues presented monkeys with two successive objects at distinct spatial positions, each of which was built from a set of basic elements (squares). The objects differed only by a single square, and the task was to discriminate which square was missing. The authors observed neurons in parietal area 7a that coded for the spatial relationship between the missing square and the probe object, and that were invariant to overall spatial position. This suggests that parietal neurons code explicitly for the spatial relations among object parts (Chafee et al., 2007).

5. Scalar codes for arbitrary categories

However, cells in posterior parietal cortex also seem to be sensitive to "what" as well as to "where". For example, when a monkey learns to classify shapes (Fitzgerald et al., 2011) or motion directions (Freedman and Assad, 2006) into arbitrary groupings, LIP neurons abstract over the physical identity of objects or stimulus features and instead code for the category itself (Freedman and Assad, 2016). Interestingly, there is evidence that the patterns of categorical coding that emerge can be remarkably low-dimensional. Intuitively, after the monkey has been trained to group shapes into categories A and B one might expect roughly equal average firing rates among neurons coding for the two classes. However, within a single animal, neural populations show globally heightened firing for (an arbitrary) one of the two categories, so that class membership can be decoded along a single dimension indicated by average firing rate (Fitzgerald et al., 2013). A close examination of the distribution of selectivity that emerges from the well-studied random dot motion paradigm reveals a similar phenomenon. When the monkey is exposed to dot motion patterns that move either right or left with high or low coherence, input-selective LIP neurons seem to monolithically prefer high/right > low/right > low/left > high/left. Again, these finding seem to indicate that LIP has compressed the high-dimensional manifold on which the physical stimuli lie into a scalar code that signals "rightness" vs. "leftness" via a one-dimensional axis of neural responsivity (Chafee, 2013). A similarly idiosyncratic neural bias for one attended feature over another has been observed in fMRI signals from a cohort of human participants judging visual stimuli according to their colour, motion direction, or identity, with maximal bias in the dorsal stream (Gong and Liu, 2019). A model that explains these phenomena also accounts for the one-dimensional linkage between delay-period and spontaneous activity in LIP during a spatial oculomotor task (Ganguli et al., 2008).

Stepping back, one interpretation of these observations is that the primate brain has evolved a distinctive computational motif, whereby high-dimensional inputs are projected onto a low-dimensional manifold in posterior parietal cortex, and that this manifold encodes their relative position in physical space (in the case of spatially selective neurons) or in an abstract space (in the case of category-selective neurons)

A

B

○ Hippocampus ● Entorhinal cortex ● Posterior parietal cortex

Place cells   **Grid cells**

Figure from Jeffrey & Burgess (2006)

O'Keefe & Dostrovsky (1971)   *Hafting et al. (2005)*   *Viswanathan & Nieder (2013)*   *Roitman et al. (2007)*
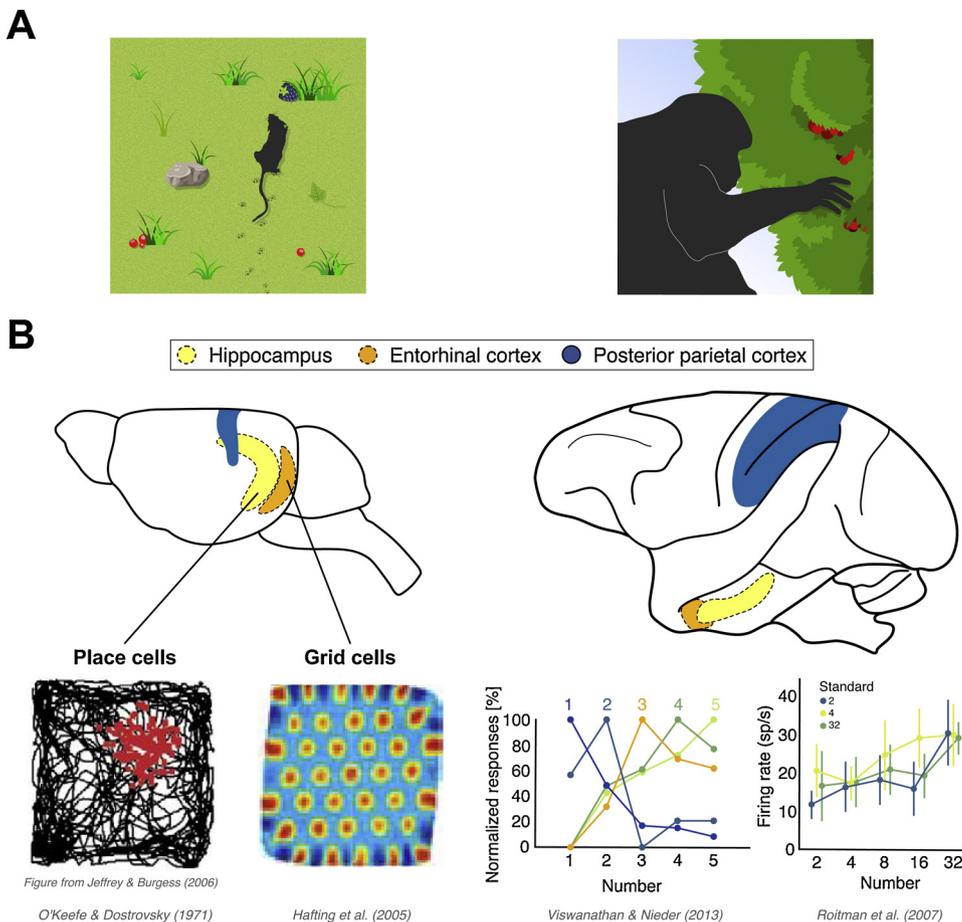
Fig. 2. **Representations of space in rodents and primates.** A. Rodents spend much of their waking lives engaging in locomotion, which requires an allocentric representation of space (left). Primates, by contrast, spend more time than rodents handling objects in peripersonal space, such as when foraging for fruit (right). B. This may be reflected in the relative size and functional significance of the hippocampal-entorhinal system (yellow/orange) in the rodent, and the posterior parietal cortex (blue) in the monkey. The hippocampus and entorhinal cortex house place cells (O'Keefe and Dostrovsky, 1971) [Reprinted from *Trends in cognitive sciences*, 10(1), Jeffery, K. J., & Burgess, N., 'A metric for the cognitive map: found at last?', 1–3, 2006, with permission from Elsevier] and grid cells (Hafting et al., 2005) [Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Nature, Microstructure of a spatial map in the entorhinal cortex, Hafting, T. et al., 2005], which have been argued to code for allocentric position and the structure of allocentric space itself. The posterior parietal cortex contains neurons sensitive to position on a line, either in physical space or in an abstract space representing magnitude or number (Viswanathan and Nieder, 2013), as well as cells suited for coding the structure of the line itself (Roitman et al., 2007) (licensed under CC-BY).

and may allow translation between the two. Low dimensional representations sacrifice coding efficiency for robustness and are thus particularly useful for strong transfer to novel settings –the functionality that is limited in contemporary deep networks trained to label images. Thus, one parsimonious description of the division of labour between dorsal and ventral streams for object and scene perception appeals to the dimensionality of neural representations. We speculate that ventral stream structures form efficient, high-dimensional codes for object identity, while dorsal stream structures learn low-dimensional representations of relational structure that are optimised for abstraction and generalisation.

6. Number and magnitude

The cases highlighted above imply that neural coding of categories in LIP is approximately one-dimensional. A one-dimensional manifold is a line, and humans use the concept of *magnitude* to describe the relative position that stimuli occupy on a line, especially in the case where high-intensity and low-intensity stimuli fall at opposite ends of the continuum (a "prothetic" representation). Humans have developed a symbolic system for quantifying and communicating magnitude, denoted by numbers, and the parietal cortex is strongly associated with numerical cognition. Firstly, there is evidence that parietal damage and disruption can lead to failures of counting and dyscalculia in humans (Cohen Kadosh et al., 2007; Martory et al., 2003). Secondly, studies using non-symbolic number also suggest that single cells in macaque LIP code for numerosity. For example, when performing a number matching task, individual cells code for distinct cardinalities with approximately Gaussian tuning curves (Nieder and Miller, 2003). This tuning is preserved across different input modalities (number of dots or sounds), indicating that the neural code for number in LIP abstracts

over physical stimulus properties (Nieder, 2012). In a different task that involves comparing whether the number of dots in an array is greater or less than a reference, LIP cells code for number with a rate (or inverse rate) code – individual neurons fire more frequently to larger (or smaller) numbers (Roitman et al., 2007). These findings have some analogues in human data. For example, when human participants report confidence in their memories on a numerical scale, a monotonic code for response number is observed in single cells in human PPC (Rutishauser et al., 2018) (Fig. 4a). Moreover, magnetoencephalographic (MEG) signals over human parietal cortex monotonically code the pulsatile rate of an audio-visual stimulus on a low-dimensional axis (Cao et al., 2019) (Fig. 4b).

Posterior parietal cortex thus seems to contain two distinct codes for magnitude: a heterogenous code (e.g. for specific numbers) and a "summation" code in which firing rate increases or decreases as number or magnitude grows. These signals would be ideally suited to respectively relay the place occupied by a given cardinality on the magnitude line (via a heterogenous code) and the structure of the line itself (via a summation code) just as place and grid cells may code for distinct physical loci and overall geometry of allocentric space (Behrens et al., 2018) (Fig. 2b, lower panels). These abstract codes may even be useful for structuring behaviour in time. For example, LIP cells code heterogeneously for the number of movements elapsed in a structured sequence of push and turn movements (Sawamura et al., 2002), akin to newly-discovered hippocampal cells that keep track of the number of laps that a rat has run around a track (Sun et al., 2019).

We thus appeal to the connection between neural coding for number and scalar codes for objects and categories in posterior parietal cortex. It is as if PPC represents items in a relational space, with otherwise symmetric classes (e.g. A and B) being arbitrarily coded as if one were "more" and the other were "less" (Chafee, 2013). Indeed, if one
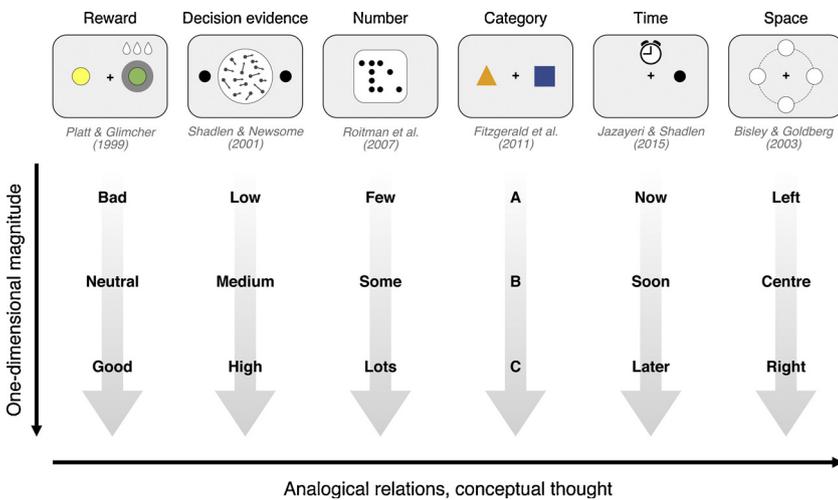
Fig. 3. A theory of magnitude. We reprise and extend ATOM theory (Walsh, 2003) which argues that overlapping neural codes for space, time and number in the parietal cortex are all manifestations of a common representation of prothetic magnitude. We suggest that other quantities that are coded monotonically in LIP cells – including value, decision information, and even arbitrary categories – may reflect the projection of high-dimensional signals onto a one-dimensional (or low-dimensional) manifold that is suited for generalisation and transfer, i.e. for the understanding of the analogical relations between abstract concepts. Here, we emphasise those concepts that share one-dimensional structure.
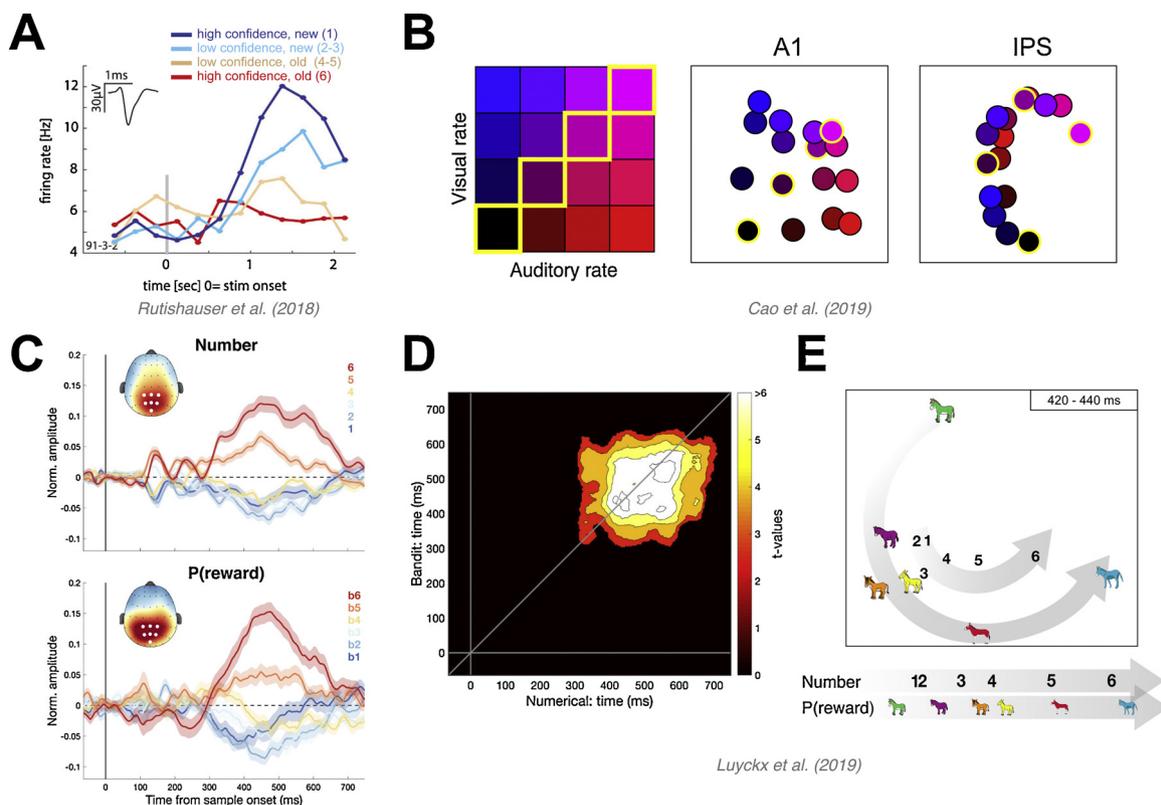
Fig. 4. Neural evidence for low-dimensional coding in human posterior parietal cortex. A. Example single neuron in human PPC whose firing rate depends on memory strength as reported on a numerical scale, with highest response for number "6" and lowest for number "1" (Rutishauser et al., 2018). (Redistributed under Elsevier user licence) B. Whilst undergoing MEG recordings, humans judged the rate of a multisensory stimulus that pulsated simultaneously and independently in the visual (flashes) and auditory (flutter) domains. Left panel: colours show different auditory and visual rate conditions. Right panel: multidimensional scaling (MDS) reveals that brain signals over PPC represented fused multisensory rate with a low-dimensional signal (Cao et al., 2019), whereas A1 coded for the input rate. C. Univariate EEG signals evoked by symbolic numbers (1–6; top panel) and visual images that paid out reward with fixed probability (ranked by their value, b1-b6; bottom panel). D. Cross-validated representational similarity analysis shows that common coding for numbers and bandits with the same magnitude occurred at 300–700 ms post-stimulus. E. MDS shows the low-dimensional form of the shared neural representation. Data in C–E from (Luyckx et al., 2019) (licensed under CC-BY).

considers the wider range of variables that seem to be encoded by LIP neurons, including time (Jazayeri and Shadlen, 2015), value (Platt and Glimcher, 1999) and decision evidence (Shadlen and Newsome, 2001), as well as number (Nieder, 2016), these variables all share the common property that they can be represented on a single dimension (soon vs. later; good vs. bad; strong vs. weak; few vs. many), suggestive of a domain-general low-dimensional coding scheme in the primate parietal cortex (Fig. 3). As alluded to above, it has been proposed that the spatial coding properties of cells in the rodent hippocampal-entorhinal system are used as basis functions for forms of nonspatial cognition that take place in abstract metric spaces. Here, we suggest an equivalent scheme in the primate parietal cortex that relies on cells sensitive to locations in egocentric rather than allocentric space.

Low-dimensional codes can facilitate generalisation between otherwise incommensurable domains. Indeed, cognitive scientists have long noted human behavioural phenomena that point to automatic
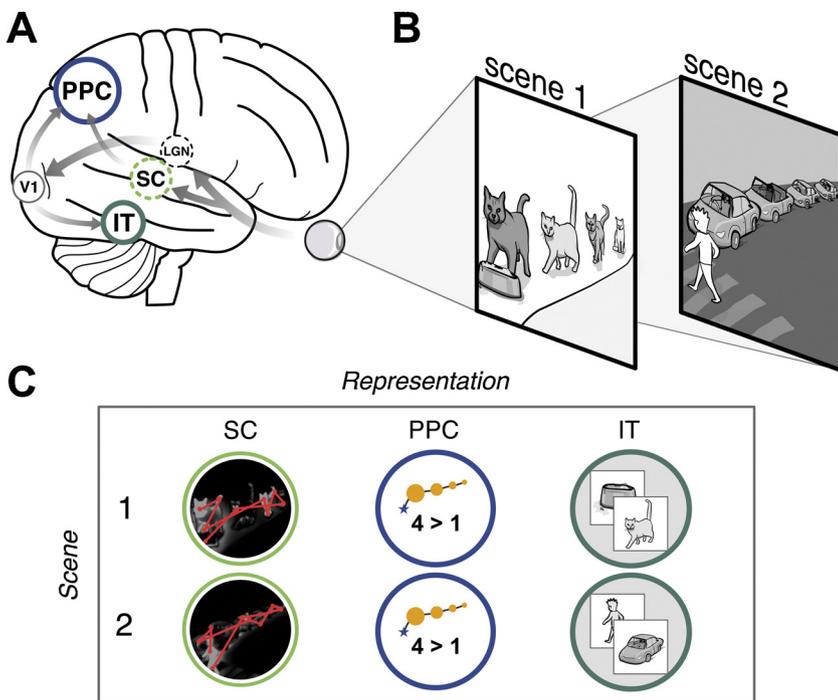
Fig. 5. Proposed neural circuit for structure learning and generalisation. A. The posterior parietal cortex (PPC) not only receives inputs from early visual cortex (V1), but also has rapid access to information about the relative salience of different spatial locations, via inputs from the superior colliculus (SC). This salience map defines the "gaze affordances" for the scene, i.e. the likely transitions between scene locations that will be made with the eyes. B. Humans readily understand the analogy between the two images shown on the left and right, whereby cats are queuing for food and commuters are queuing in a traffic jam. C. We propose that signals concerning likely saccadic targets in a scene (from SC) convey information about object relations to parietal cortex. This affords parietal cortex the opportunity to learn invariances over scene structure, which for example facilitate generalisation between the images in panel B. Neural codes for scene structure (in parietal cortex) and object identity (in inferotemporal cortex) may then be combined into a comprehensive scene representation in downstream structures, such as the medial temporal lobe or retrosplenial cortex.

generalisation between space, time and number, leading to the proposal that the parietal cortex houses a domain-general magnitude representation (Walsh, 2003). For example, in the well-known SNARC (spatial-numerical association of response codes) effect, participants from Western cultures are faster at responding to larger numbers on the right side of visual space, as if their mental number line recycled a spatial representation arranged from left (low numbers) to right (high numbers), consistent with their learned reading direction (Dehaene et al., 1993). A plethora of other biases suggest deep homologies in the way that space, time and number are coded (Hubbard et al., 2005; Lourenco and Longo, 2011). In fact, the idea that space provides a basis for more abstract form of inference is an old one in psychology, dating back to the debate between analogue vs. propositional representations in the 1980s (Pylyshlyn, 1981).

One recent paper set out to directly test one of these homologies in human brain signals (Luyckx et al., 2019). A precursor to this research was the demonstration that when humans view Arabic digits, patterns of EEG activity vary smoothly with numerical distance, such that the multivariate signal evoked by "3" is more similar to that elicited by "4" than "5" (Spitzer et al., 2017; Teichmann et al., 2018). Because the numbers shown were symbolic, this signal cannot be explained by sensory features of the stimulus and thus must reflect an abstract code for numerical magnitude. Luyckx and colleagues (Fig. 4c–e) replicated this finding, but additionally asked participants to choose among pairs of 6 possible visual images (animal pictures) on the basis of their payout probability (Luyckx et al., 2019). After training on this "bandit" task, the multivariate EEG signals for numbers 1–6 predicted those evoked by the animal pictures, ranked by their reward probability, with the highest-valued bandit eliciting homologous patterns to number "6" and the lowest to number "1". This occurred even though no numerical information had ever been associated with the visual images, suggesting that the neural representation of the bandits was structured by the pre-existing representation of numerical magnitude in the brains of the participants. Interestingly, the multivariate EEG signal for both numbers and bandits was partly (but not wholly) driven by a univariate (i.e. one-dimensional) summation code with a maximum over centro-parietal electrodes, consistent with the theory of parietal function advanced here. In related work, others have noted overlapping representations in parietal BOLD signals for one-dimensional metrics in

distinct domains, such as spatial, temporal and social distance (Parkinson et al., 2014). However, it has been debated whether these results imply a common representation or simply shared resources in the parietal cortex (Borghesani et al., 2019).

7. How is structure learned from experience?

Thus far, we have described a theory of the parietal cortex that emphasises its role in representing the relational structure of objects and categories, in both physical and abstract metric spaces. Specifically, we suggest that parietal neurons tuned for spatial position are recycled for making relational inferences in nonspatial domains. The theory appeals to the seemingly abstract and low-dimensional codes that are observed in neural populations in LIP. However, we have not explained how these coding properties may be acquired by experience. In fact, building a neural information processing system that learns about abstract relational structure remains a grand challenge for both neuroscientists and AI researchers. Nevertheless, in the final part of this article, we advance some ideas which we hope may shed light on this question.

How can a neuron learn to explicitly represent the structure of the natural world in a way that is invariant to its contents? The challenge is that learning of invariant structure must be grounded in experience, and that experience is dominated by the contents of sensory inputs. Consider the case of an agent learning an abstract representation of number. Confronted with an array of three wholly novel objects, a human who is unable to identify the objects would have no difficulty counting them, because she has learned the concept of "three". But how is this concept acquired in the first place?

One possibility is that passive exposure to the contents of sensory experience itself is sufficient to ground structure learning. This is a popular view among machine learning researchers, who hope that relational abstractions will emerge in neural networks that are built to be sufficiently expressive (being endowed with layer-wise depth and millions of parameters) and trained on a broad enough distribution of static images or videos. In fact, there is recent evidence that codes resembling those for numerosity emerge spontaneously in the deep layers of a neural network trained to label objects from a very large database (Nasr et al., 2019). However, humans and monkeys who exhibit neural

representations of number do not have the luxury of learning via millions of labelled examples, and so it is unclear whether this approach is ultimately practicable for biological systems. Recent advances in unsupervised methods have led to the hope that human-interpretable concepts can be disentangled without labels and from pixels alone (Higgins et al., 2017), but learning abstractions that are not tied to physical sensory properties (e.g. Fig. 5) remains an elusive goal in AI research.

Here, we draw upon an alternative view: that structure learning is ultimately grounded in action. In fact, the ability to take actions is arguably a necessary precursor to learning about space itself. This is most evident in the case of allocentric space, whose structure is disclosed by actively experiencing the state transitions that occur during navigation (at least for mammals that never experience a bird's eye view of the environment). However, it is almost certainly the case that the structure of egocentric space is learned by reaching and grasping for objects and fixating the objects with the eyes at various spatial positions and in different planes of depth. To cycle back to our critique of deep learning as a computational theory of visual scene understanding, this argument entails that objectness cannot be learned without a control policy that allows objects to be picked up and handled, and scene structure cannot be understood without a shifting gaze that explores multiple different scene elements in turn.

We thus argue that information about viable or probable actions is potentially an important source of *input* to any neural system that is built to learn about the structure of natural scenes. This idea draws upon the longstanding theme, dating back to theories of "direct perception" (Gibson, 1979), whereby objects are understood by the actions that they afford. In fact, it is salient that PPC neurons code for spatial information in a mixed frame of reference that bridges sensory input and motor output (Bennur and Gold, 2011). For example, microstimulation of LIP neurons elicits a saccade towards their preferred location, whereas cells in the more anterior AIP region become active during reaching towards an object in peri-personal space (Andersen and Cui, 2009). In other words, these parietal areas putatively coding for relations among objects multiplex sensory inputs with signals pertaining to the motor actions required to handle or fixate them. Indeed, this would explain why the neural build-up signals that are observed in advance of oculomotor behaviour do not seem to be causally necessary for saccadic decisions in simple discrimination tasks (Katz et al., 2016) but are important for more abstract category decisions (Zhou and Freedman, 2019). Another framing of this claim is that when we talk about "affordances" we are referring to a transition probability matrix for peri-personal space, describing the probability that an object will be a future goal of reaching or target of fixation. In the primate, this transition matrix may be encoded in parallel with that for allocentric space in the MTL, with the two structures offering complementary coding schemes for the encoding of abstract concepts.

Actions are typically thought of as an output from, rather than an input to, sensorimotor systems. However, at least in the case of eye movements, the circuitry of the oculomotor in principle allows information about candidate actions to arrive to posterior parietal cortex contiguously with (or even before) sensory inputs (White et al., 2017b). In parallel to the classic thalamocortical pathway, visual information is routed to the primate superior colliculus (SC) where candidate saccades are computed rapidly in a viewer-centred frame of reference. It is widely accepted that SC codes candidate target spatial positions for saccades on the basis of their bottom-up salience and top-down relevance (White et al., 2017a). Visuomotor neurons signalling this information in SC fire within 50–60 ms of stimulus onset, and tracing studies reveal that IPL is a major target for SC outputs (Clower et al., 2001). It is possible, thus, that posterior parietal cortex receives information about the likely saccadic eye movements that will be taken towards a visual scene (Fig. 5). Critically for our argument, these signals will inevitably carry information about the relational geometry of the objects that are present in the scene, stripped of information about

the object identity. This provides an opportunity for IPL to form representations of scene structure independent of scene contents, and thus potentially to support relational generalisation between scenes with comparable structure. This theory is currently a speculation, but it leads to a number of testable predictions both for neuroscience and machine learning research.

8. Outlook

In this article, we have attempted to join the dots between different literatures: deep learning models as computational theory of representational learning, the neuropsychology and neurophysiology of the parietal cortex, and the cognitive origins of abstract thought. However, there are parts of the puzzle that we struggle to incorporate into our theory. In closing, we highlight some of the most salient.

Firstly, we offer a speculation about the dimensionality of neural signals in the primate dorsal and ventral streams: we argue that sparse neural codes in ventral stream support efficient object recognition, whereas low-dimensional codes in the dorsal stream permit generalisation of categories and concepts. Whilst there is preliminary evidence in favour of this view (Fitzgerald et al., 2013; Quiroga et al., 2008), it remains a speculation. Indeed, others have suggested that IT neurons may encode the principal generative factors in datasets composed of faces (Chang and Tsao, 2017) or abstract shapes (Op de Beeck et al., 2001), suggestive of the sort of representation we associated with the dorsal stream. We also acknowledge the complexity of the brain structures we discuss and their implication in a broad range of functions; we have focussed on the role of the posterior parietal cortex in scene understanding, but our theory does not (for example) encompass its participation in a broad range of attentional and mnemonic functions.

Secondly, whilst there is good evidence for a dissociation between two frames of reference – a representation of allocentric space in the medial temporal lobe and egocentric space in the parietal cortex – this may be a simplification. In the primate, hippocampal neurons are sensitive to spatial views of a scene in a partly egocentric frame of reference (Rolls and O'Mara, 1995). Even in the rodent, cells in the hippocampal formation do not code exclusively for space – they are also sensitive to objects (Hoydal et al., 2019), landmarks (Manns and Eichenbaum, 2009) and boundaries (O'Keefe and Burgess, 1996). Moreover, hippocampal lesions impair transitive inference in rodents (Dusek and Eichenbaum, 1997), and BOLD signals in the hippocampus code for other abstract relations that occur naturally in one dimension, such as position on a social dominance hierarchy (Kumaran et al., 2012). This suggests that the monotonic coding principles associated here with the parietal cortex might also find expression in the medial temporal lobe.

Relatedly, grid cells in the entorhinal cortex of humans (Nau et al., 2018) and nonhuman primates (Killian et al., 2012) show six fold modulation by the trajectory of saccadic eye movements across a visual scene. This clearly blurs the distinction we have proposed between coding of allocentric space in the medial temporal lobe, and peri-personal space in the parietal cortex. In fact, recent studies that measure hexagonal symmetry in human BOLD signals have observed "grid-like" coding of stimuli with monotonically varying object features (Constantinescu et al., 2016) or for modalities that lack overt spatial connotation, such as odours (Bao et al., 2019). We find these results puzzling, because it seems like a more natural coding scheme for these spaces would factorise the stimuli according to the relevant axes (e.g. neck and leg length of a bird, or concentration of pine vs. banana odour) via the sort of monotonic coding scheme proposed here. We think it is likely that brain regions will differ with respect to the extent that they exhibit quadrilateral code (i.e. with one dimension aligned to each axis) and a hexagonal code, as exhibited by grid cells, but this remains for future research to discern. In sum, thus, whilst none of these findings invalidate the claims made here, more research is needed to

delineate the precise division of labour between the medial temporal lobe and parietal systems for representing abstract relational information in brains of rodents and monkeys.

Finally, the distinction we draw here between parietal cortex and MTL also overlooks a literature that has studied how these regions interact. For example, neural information from the caudal intraparietal area is routed to the hippocampus both directly, and indirectly via the posterior cingulate cortex and retrosplenial cortex (Kravitz et al., 2011). Computational theories propose how this pathway might subserve the conversion of spatial information from an egocentric to an allocentric frame of reference (Byrne et al., 2007). Integrating these ideas into the theory proposed here is beyond the scope of the current review, but would be of considerable interest for future work.

We began by noting that adult humans are capable of learning to represent a rich set of concepts. Many of the more abstract concepts that we understand and communicate are fundamentally relational in nature. For example, the concept of "friendship" implies a social relation between two individuals, and the concept of "commerce" refers to a relational system by which assets flow within an economy. We think it is likely that learning about these complex relational concepts is grounded on an understanding of how physical space is organised, which in turn arises when we make actions in that space. The deep connection between space and abstract concepts is also reflected in the metaphors and similes that we use in natural language (Wolff and Gentner, 2011). For example, terms invoking the Gestalt principles of good continuity, containment or scale are used to denote abstract relations in nonspatial domains, as when we state that two individuals' views are "aligned" or that a group member is an "insider". In other words, learning about the organisation of space may be a necessary precursor to forming a rich and complex model of the world.

Funding

Acknowledgements

Appendix A. The Peer Review Overview and Supplementary data

The Peer Review Overview and Supplementary data associated with this article can be found in the online version, at doi:https://doi.org/10.1016/j.pneurobio.2019.101717.

References

Andersen, R.A., Cui, H., 2009. Intention, action planning, and decision making in parietal-frontal circuits. Neuron 63, 568–583.

Aragao Rda, S., Rodrigues, M.A., de Barros, K.M., Silva, S.R., Toscano, A.E., de Souza, R.E., Manhaes-de-Castro, R., 2011. Automatic system for analysis of locomotor activity in rodents–a reproducibility study. J. Neurosci. Methods 195, 216–221.

Bao, X., Gjorgieva, E., Shanahan, L.K., Howard, J.D., Kahnt, T., Gottfried, J.A., 2019. Grid-like neural representations support olfactory navigation of a two-dimensional odor space. Neuron 102 1066–1075 e1065.

Barrett, D.G.T., Hill, F., Santoro, A., Morcos, A.S., Lillicrap, T., 2018. Measuring Abstract Reasoning in Neural Networks. arXiv:1807.04225v1. .

Behrens, T.E.J., Muller, T.H., Whittington, J.C.R., Mark, S., Baram, A.B., Stachenfeld, K.L., Kurth-Nelson, Z., 2018. What is a cognitive map? Organizing knowledge for flexible behavior. Neuron 100, 490–509.

Bellmund, J.L.S., Gardenfors, P., Moser, E.I., Doeller, C.F., 2018. Navigating cognition: spatial codes for human thinking. Science 362.

Bennur, S., Gold, J.I., 2011. Distinct representations of a perceptual decision and the associated oculomotor plan in the monkey lateral intraparietal area. J. Neurosci. 31, 913–921.

Black, F.W., Strub, R.L., 1976. Constructional apraxia in patients with discrete missile wounds of the brain. Cortex 12, 212–220.

Borghesani, V., de Hevia, M.D., Viarouge, A., Pinheiro-Chagas, P., Eger, E., Piazza, M., 2019. Processing number and length in the parietal cortex: sharing resources, not a common code. Cortex 114, 17–27.

Brendel, W., Rauber, J., Bethge, M., 2017. Decision-Based Adversarial Attacks: Reliable Attacks Against Black-Box Machine Learning Models. arXiv:1712.04248. .

Burgess, C.P., Matthey, L., Watters, N., Kabra, R., Higgins, I., Botvinick, M., Lerchner, A., 2019. MONet: Unsupervised Scene Decomposition and Representation. arXiv:1901.11390. .

Byrne, P., Becker, S., Burgess, N., 2007. Remembering the past and imagining the future: a neural model of spatial memory and imagery. Psychol. Rev. 114, 340–375.

Cao, Y., Summerfield, C., Park, H., Giordano, B.L., Kayser, C., 2019. Causal inference in the multisensory brain. Neuron 102 1076-1087 e1078.

Chafee, M.V., 2013. A scalar neural code for categories in parietal cortex: representing cognitive variables as "more" or "less". Neuron 77, 7–9.

Chafee, M.V., Averbeck, B.B., Crowe, D.A., 2007. Representing spatial relationships in posterior parietal cortex: single neurons code object-referenced position. Cereb. Cortex 17, 2914–2932.

Chang, L., Tsao, D.Y., 2017. The code for facial identity in the primate brain. Cell 169 1013-1028 e1014.

Chechlacz, M., Humphreys, G.W., 2014. The enigma of Balint's syndrome: neural substrates and cognitive deficits. Front. Hum. Neurosci. 8, 123.

Clower, D.M., West, R.A., Lynch, J.C., Strick, P.L., 2001. The inferior parietal lobule is the target of output from the superior colliculus, hippocampus, and cerebellum. J. Neurosci. 21, 6283–6291.

Cohen Kadosh, R., Cohen Kadosh, K., Schuhmann, T., Kaas, A., Goebel, R., Henik, A., Sack, A.T., 2007. Virtual dyscalculia induced by parietal-lobe TMS impairs automatic magnitude processing. Curr. Biol. 17, 689–693.

Constantinescu, A.O., O'Reilly, J.X., Behrens, T.E.J., 2016. Organizing conceptual knowledge in humans with a gridlike code. Science 352, 1464–1468.

Dehaene, S., Bossini, S., Giraux, P., 1993. The mental representation of parity and number magnitude. J. Exp. Psychol. Gen. 122, 371–396.

Dordek, Y., Soudry, D., Meir, R., Derdikman, D., 2016. Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis. Elife 5, e10094.

Doumas, L.A., Hummel, J.E., Sandhofer, C.M., 2008. A theory of the discovery and predication of relational concepts. Psychol. Rev. 115, 1–43.

Dusek, J.A., Eichenbaum, H., 1997. The hippocampus and memory for orderly stimulus relations. Proc. Natl. Acad. Sci. U. S. A. 94, 7109–7114.

Fitzgerald, J.K., Freedman, D.J., Assad, J.A., 2011. Generalized associative representations in parietal cortex. Nat. Neurosci. 14, 1075–1079.

Fitzgerald, J.K., Freedman, D.J., Fanini, A., Bennur, S., Gold, J.I., Assad, J.A., 2013. Biased associative representations in parietal cortex. Neuron 77, 180–191.

Fleuret, F., Li, T., Dubout, C., Wampler, E.K., Yantis, S., Geman, D., 2011. Comparing machines and humans on a visual categorization test. Proc. Natl. Acad. Sci. U. S. A. 108, 17621–17625.

Freedman, D.J., Assad, J.A., 2006. Experience-dependent representation of visual categories in parietal cortex. Nature 443, 85–88.

Freedman, D.J., Assad, J.A., 2016. Neuronal mechanisms of visual categorization: an abstract view on decision making. Annu. Rev. Neurosci. 39, 129–147.

Freedman, D.J., Ibos, G., 2018. An integrative framework for sensory, motor, and cognitive functions of the posterior parietal cortex. Neuron 97, 1219–1234.

Friedman-Hill, S.R., Robertson, L.C., Treisman, A., 1995. Parietal contributions to visual feature binding: evidence from a patient with bilateral lesions. Science 269, 853–855.

Ganguli, S., Bisley, J.W., Roitman, J.D., Shadlen, M.N., Goldberg, M.E., Miller, K.D., 2008. One-dimensional dynamics of attention and decision making in LIP. Neuron 58, 15–25.

Geirhos, R., Medina Temme, C.R., Rauber, J., Schutt, H.H., Bethge, M., Wichmann, F., 2018. Generalisation in Humans and Deep Neural Networks. arXiv:1808.08750. .

Gentner, D., 1983. Structure mapping: a theoretical framework for analogy. Cogn. Sci. 7, 155–170.

Gibson, J.J., 1979. The Ecological Approach to Visual Perception. Houghton Mifflin, Boston.

Gold, J.I., Shadlen, M.N., 2007. The neural basis of decision making. Annu. Rev. Neurosci. 30, 535–574.

Gong, M., Liu, T., 2019. Biased Neural Coding of Feature-Based Attention in Human Brain. https://www.biorxiv.org/content/10.1101/688226v1.

Goodale, M.A., Milner, A.D., 1992. Separate visual pathways for perception and action. Trends Neurosci. 15, 20–25.

Hafting, T., Fyhn, M., Molden, S., Moser, M.B., Moser, E.I., 2005. Microstructure of a spatial map in the entorhinal cortex. Nature 436, 801–806.

Hassabis, D., Kumaran, D., Summerfield, C., Botvinick, M., 2017. Neuroscience-inspired artificial intelligence. Neuron 95, 245–258.

Higgins, I., Sonnerat, N., Matthey, L., Pal, A., Burgess, C.P., Bosnjak, M., Shanahan, M., Botvinick, M., Hassabis, D., 2017. SCAN: Learning Hierarchical Compositional Visual Concepts. arXiv:1707.03389. .

Hoydal, O.A., Skytoen, E.R., Andersson, S.O., Moser, M.B., Moser, E.I., 2019. Object-vector coding in the medial entorhinal cortex. Nature 568, 400–404.

Hubbard, E.M., Piazza, M., Pinel, P., Dehaene, S., 2005. Interactions between number and space in parietal cortex. Nat. Rev. Neurosci. 6, 435–448.

Jazayeri, M., Shadlen, M.N., 2015. A neural mechanism for sensing and reproducing a time interval. Curr. Biol. 25, 2599–2609.

Katz, L.N., Yates, J.L., Pillow, J.W., Huk, A.C., 2016. Dissociated functional significance of decision-related activity in the primate dorsal stream. Nature 535, 285–288.

Khaligh-Razavi, S.M., Kriegeskorte, N., 2014. Deep supervised, but not unsupervised, models may explain IT cortical representation. PLoS Comput. Biol. 10, e1003915.

Killian, N.J., Jutras, M.J., Buffalo, E.A., 2012. A map of visual space in the primate entorhinal cortex. Nature 491, 761–764.

Kravitz, D.J., Saleem, K.S., Baker, C.I., Mishkin, M., 2011. A new neural framework for visuospatial processing. Nat. Rev. Neurosci. 12, 217–230.

Kriegeskorte, N., 2015. Deep neural networks: a new framework for modeling biological vision and brain information processing. Annu. Rev. Vis. Sci. 1, 417–446.

Krizhevsky, A., Sutskever, I., Hinton, G., 2012. ImageNet classification with deep convolutional neural networks. Neural Information Processing Systems.

Kumaran, D., Melo, H.L., Duzel, E., 2012. The emergence and representation of knowledge about social and nonsocial hierarchies. Neuron 76, 653–666.

Lake, B.M., Salakhutdinov, R., Tenenbaum, J.B., 2015. Human-level concept learning through probabilistic program induction. Science 350, 1332–1338.

Lake, B.M., Ullman, T.D., Tenenbaum, J.B., Gershman, S.J., 2016. Building machines that learn and think like people. CBMM Memo 046.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Leibo, J.Z., de Masson d'Autume, C., Zoran, D., Amos, D., Beattie, C., Anderson, K., García Castañeda, A., Sanchez, M., Green, S., Gruslys, A., Legg, S., Hassabis, D., Botvinick, M., 2018. Psychlab: A Psychology Laboratory for Deep Reinforcement Learning Agents. arXiv:1801.08116. .

Lourenco, S.F., Longo, M.R., 2011. Origins and development of generalized magnitude representation. In: Dehaene, S., Brannon, E.M. (Eds.), Space, Time and Number in the Brain: Searching for the Foundations of Mathematical Thought. Elsevier Academic Press, San Diego, CA, pp. 225–244.

Luyckx, F., Nili, H., Spitzer, B., Summerfield, C., 2019. Neural structure mapping in human probabilistic reward learning. Elife 8.

Manns, J.R., Eichenbaum, H., 2009. A cognitive map for object memory in the hippocampus. Learn. Mem. 16, 616–624.

Marblestone, A.H., Wayne, G., Kording, K.P., 2016. Toward an integration of deep learning and neuroscience. Front. Comput. Neurosci. 10, 94.

Marcus, G., 2018. Deep Learning: A Critical Appraisal. arXiv:1801.00631. .

Martory, M.D., Mayer, E., Pegna, A.J., Annoni, J.M., Landis, T., Khateb, A., 2003. Pure global acalculia following a left subangular lesion. Neurocase 9, 319–328.

Murphy, G., 2002. The Big Book of Concepts. MIT Press, Cambridge, MA.

Nasr, K., Viswanathan, P., Nieder, A., 2019. Number detectors spontaneously emerge in a deep neural network designed for visual object recognition. Sci. Adv. 5, eaav7903.

Nau, M., Navarro Schroder, T., Bellmund, J.L.S., Doeller, C.F., 2018. Hexadirectional coding of visual space in human entorhinal cortex. Nat. Neurosci. 21, 188–190.

Nguyen, A., Yoshinski, J., Clune, J., 2014. Deep Neural Networks Are Easily Fooled: High Confidence Predictions for Unrecognizable Images. arXiv:1412.1897. .

Nieder, A., 2012. Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. Proc. Natl. Acad. Sci. U. S. A. 109, 11860–11865.

Nieder, A., 2016. The neuronal code for number. Nat. Rev. Neurosci. 17, 366–382.

Nieder, A., Miller, E.K., 2003. Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. Neuron 37, 149–157.

O'Keefe, J., Burgess, N., 1996. Geometric determinants of the place fields of hippocampal neurons. Nature 381, 425–428.

O'Keefe, J., Dostrovsky, J., 1971. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res. 34, 171–175.

Op de Beeck, H., Wagemans, J., Vogels, R., 2001. Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. Nat. Neurosci. 4, 1244–1252.

Parkinson, C., Liu, S., Wheatley, T., 2014. A common cortical metric for spatial, temporal, and social distance. J. Neurosci. 34, 1979–1987.

Passingham, R.E., Wise, S.P., 2012. The Neurobiology of the Prefrontal Cortex. Oxford University Press, Oxford.

Platt, M.L., Glimcher, P.W., 1999. Neural correlates of decision variables in parietal cortex. Nature 400, 233–238.

Pylyshyln, Z., 1981. The imagery debate: analogue media versus tacit knowledge. Psychol. Rev. 88, 16–45.

Quiroga, R.Q., Kreiman, G., Koch, C., Fried, I., 2008. Sparse but not' grandmother-cell' coding in the medial temporal lobe. Trends Cogn. Sci. 12, 87–91.

Roitman, J.D., Brannon, E.M., Platt, M.L., 2007. Monotonic coding of numerosity in macaque lateral intraparietal area. PLoS Biol. 5, e208.

Rolls, E.T., O'Mara, S.M., 1995. View-responsive neurons in the primate hippocampal complex. Hippocampus 5, 409–424.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Kholsa, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. arXiv:1409.0575. .

Rutishauser, U., Aflalo, T., Rosario, E.R., Pouratian, N., Andersen, R.A., 2018. Single-neuron representation of memory strength and recognition confidence in left human posterior parietal cortex. Neuron 97 209–220 e203.

Santoro, A., Raposo, D., Barrett, D.G.T., Malinowski, M., Pascanu, R., Battaglia, P., Lillicrap, T., 2017. A Simple Neural Network Module for Relational Reasoning. arXiv:1706.01427. .

Sawamura, H., Shima, K., Tanji, J., 2002. Numerical representation for action in the parietal cortex of the monkey. Nature 415, 918–922.

Schank, R.C., Abelson, R.P., 1977. Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures. L. Erlbaum, Hillsdale, N.J.

Scholte, H.S., Losch, M.M., Ramakrishnan, K., de Haan, E.H.F., Bohte, S.M., 2018. Visual pathways from the perspective of cost functions and multi-task deep neural networks. Cortex 98, 249–261.

Shadlen, M.N., Newsome, W.T., 2001. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J. Neurophysiol. 86, 1916–1936.

Spitzer, B., Waschke, L., Summerfield, C., 2017. Selective overweighting of larger magnitudes during numerical comparison. Nat. Hum. Behav. 1.

Stachenfeld, K., Botvinick, M.M., Gershman, S.J., 2014. Design principles of hippocampal cognitive maps. Neural Inform. Process. Syst. 27.

Sun, C., Yang, W., Martin, J., Tonegawa, S., 2019. CA1 Pyramidal Cells Organize an Episode by Segmented and Ordered Events. https://www.biorxiv.org/content/10.1101/565689v1.

Sussman, R.W., Tattersall, I., 1981. Behavior and ecology ofmacaca fascicularis in Mauritius: a preliminary study. Primates 22, 192–205.

Teichmann, L., Grootswagers, T., Carlson, T., Rich, A.N., 2018. Decoding digits and Dice with magnetoencephalography: evidence for a shared representation of magnitude. J. Cogn. Neurosci. 30, 999–1010.

Tervo, D.G.R., Tenenbaum, J.B., Gershman, S.J., 2016. Toward the neural implementation of structure learning. Curr. Opin. Neurobiol. 37, 99–105.

Viswanathan, P., Nieder, A., 2013. Neuronal correlates of a visual "sense of number" in primate parietal and prefrontal cortices. Proc. Natl. Acad. Sci. U. S. A. 110, 11187–11192.

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J.R., van der Helm, P.A., van Leeuwen, C., 2012. A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. Psychol. Bull. 138, 1218–1252.

Walsh, V., 2003. A theory of magnitude: common cortical metrics of time, space and quantity. Trends Cogn. Sci. 7, 483–488.

White, B.J., Berg, D.J., Kan, J.Y., Marino, R.A., Itti, L., Munoz, D.P., 2017a. Superior colliculus neurons encode a visual saliency map during free viewing of natural dynamic video. Nat. Commun. 8, 14263.

White, B.J., Kan, J.Y., Levy, R., Itti, L., Munoz, D.P., 2017b. Superior colliculus encodes visual saliency before the primary visual cortex. Proc. Natl. Acad. Sci. U. S. A. 114, 9451–9456.

Wolff, P., Gentner, D., 2011. Structure-mapping in metaphor comprehension. Cogn. Sci. 35, 1456–1488.

Wu, X., Zhang, X., Shu, X., 2018. Cognitive Deficit of Deep Learning in Numerosity. arXiv:1802.05160. .

Yamins, D.L., DiCarlo, J.J., 2016. Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. 19, 356–365.

Yamins, D.L., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., DiCarlo, J.J., 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc. Natl. Acad. Sci. U. S. A. 111, 8619–8624.

Yoshinski, J., Clune, J., Nguyen, A., Fuchs, T., Lipson, H., 2015. Understanding Neural Networks Through Deep Visualization. arXiv:1506.06579. .

Zhou, Y., Freedman, D.J., 2019. Posterior parietal cortex plays a causal role in perceptual and categorical decisions. Science 365, 180–185.