Special issue: Research report

Holistic versus feature-based binding in the medial temporal lobe

CrossMark

Rebecca N. van den Honert [a,*], *Gregory McCarthy* [a,b] and
Marcia K. Johnson [a,b]

[a] *Department of Psychology, Yale University, New Haven, CT, USA*
[b] *Interdepartmental Neuroscience Program, Yale University, New Haven, CT, USA*

ARTICLE INFO

ABSTRACT

A central question for cognitive neuroscience is how feature-combinations that give rise to episodic/source memories are encoded in the brain. Although there is much evidence that the hippocampus (HIP) is involved in feature binding, and some evidence that other brain regions are as well, there is relatively little evidence about the nature of the resulting representations in different brain regions. We used multivoxel pattern analysis (MVPA) to investigate how feature combinations might be represented, contrasting two possibilities, feature-based versus holistic. Participants viewed stimuli that were composed of three source features — a person (face or body), a scene (indoor or outdoor), and an object (bike or luggage) — which were combined to make eight unique stimulus identities. We reasoned that regions that can classify the eight identities (a multiclass classification) but not the individual features (a binary classification) likely have a holistic representation of each identity. In contrast, regions that can classify the eight identities *and* can classify each feature are likely to contain feature-based representations of these identities. To further probe the extent of feature-based or holistic classification in each region, we developed and validated a novel approach that directly compares binary and multiclass classification. We found clear evidence for holistic representation in the parahippocampal cortex (PHC), consistent with theories that posit that pattern-separation-like binding mechanisms are not unique to the HIP. Further clarifying the mechanisms of feature binding should benefit from systematic comparisons of multi-feature representations and whether they vary with task, type of stimulus, and/or experience.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Episodic memory is characterized by remembering experiences as unique combinations of features, for example, what happened, where, and with whom (Johnson & Raye, 1981; Tulving, 1972; Underwood, 1969). A fundamental question for cognitive neuroscience is how such features are combined (i.e., bound) during encoding to later form the basis of the subjective experience of remembering, which entails context

(Mandler, 1980) or source (Johnson, Hashtroudi, & Lindsay, 1993) memory. Theories of episodic/source memory and medial temporal lobe (MTL) function have long ascribed the hippocampus (HIP) a central role in mnemonic binding (Cohen, Poldrack, & Eichenbaum, 1997; Davachi, 2006; Johnson & Chalfonte, 1994; O'Reilly & Rudy, 2001; Ranganath, 2010; Squire, Stark, & Clark, 2004). For example, there have been several variations of the idea that the HIP establishes associations among features that are initially represented in other brain regions [e.g., objects in perirhinal cortex (PRC) and spatial location in parahippocampal cortex (PHC); Eichenbaum, Yonelinas, & Ranganath, 2007; Howard, Kumaran, Olafsdottir, & Spiers, 2011]. However, there is increasing evidence that regions other than HIP also support mnemonic binding. Thus, recent discussions have shifted from whether the HIP uniquely subserves feature binding to what differentiates the nature of bound representations in different MTL regions, for example, the types of information that are bound, and/or the way information is bound (Cowell, Bussey, & Saksida, 2010; Norman, 2010; Shimamura, 2010).

Advances in understanding potential functional divisions of the MTL require a clearer characterization of the nature of feature binding. In approaching this task, we assume that different regions of the brain likely represent different combinations of features. For example, the fusiform face area (FFA) is thought to represent combinations of face parts (Liu, Harris, & Kanwisher, 2010; McCarthy, Puce, & Belger, 1999), the PRC to represent combinations of object parts (Erez, Cusack, Kendall, & Barense, 2016), and the PHC to represent associations among elements of a scene (Aminoff & Tarr, 2015). Rather than ask whether a region represents a particular feature or feature combination, we ask *how* particular features are combined in various regions. One possibility is that the HIP is unique in the way in which it forms multi-feature representations (Marr, 1971; O'Reilly & McClelland, 1994). Another possibility is that there are similar binding mechanisms throughout the MTL and it is the content of what is bound that varies across regions (Cowell et al., 2010; Shimamura, 2010).

Here we consider two possibilities of how feature combinations may be represented — in a *feature-based* manner or *holistically*.[1] In a feature-based representation, the feature-combination is completely predicted by the features (i.e., the whole is simply a combination of parts). If a region contains feature-based representations, it should be able to classify each feature (e.g., Persons A *vs* B, and Locations X *vs* Y) and the identity of the feature combination (e.g., AX *vs* AY *vs* BX *vs* BY). In a holistic representation, the feature-combination is represented in a way that is independent of each feature/part representation. If a region contains holistic representations, it should be able to classify the identity of a feature combination despite poor classification of the features themselves. We used simulations to validate this spectrum of feature-based versus holistic representation (Supplemental Material). Note

that one's research question will largely determine how features and feature combinations are defined; here we used combinations of people, objects, and scenes as they are common source features of episodic memories.

There is reason to believe that HIP might form holistic representations of complex feature-combinations. For example, the Complementary Learning Systems (CLS) model (Marr, 1971; McClelland & Goddard, 1996; McClelland, McNaughton, & O'Reilly, 1995; Norman, 2010; Norman & O'Reilly, 2003) proposes that incoming features are bound within the HIP such that each feature combination is represented by an independent set of neurons, yielding separate representations for similar experiences (pattern separation). The CLS model has been used to predict that representations for similar experiences should be more distinct in HIP compared to other regions (Yassa & Stark, 2011). However, an important aspect of pattern separation is that independent neurons represent each multi-feature event so that experiences are orthogonally represented (i.e., holistically represented) rather than in a way that is systematically based on features. Therefore, while evidence that pairs of stimuli are represented distinctly in the HIP (e.g., Bakker, Kirwan, Miller, & Stark, 2008; Lacy, Yassa, Stark, Muftuler, & Stark, 2011; LaRocque et al., 2013; Motley & Kirwan, 2012) is consistent with pattern separation, it does not necessarily mean that the way in which the HIP forms such distinct multi-feature representations is through pattern separation. Another possibility is that the information represented in the HIP is more distinctive because it binds representations from lower in a processing hierarchy, such as item to item or item to spatial context (Cowell et al., 2010; Shimamura, 2010). By this logic, binding in the HIP and regions lower in the hierarchy could occur via similar mechanisms, which could yield holistic or feature-based representations.

Investigations of the types of content represented in the MTL suggest the presence of holistic representation in HIP. Across studies, it appears that the HIP contains reliable representations of individual stimuli/memories but not category-level information. Specifically, MTL cortex is much better than the HIP at discriminating stimulus categories (e.g., faces, objects, scenes) (Huffman & Stark, 2014; Liang, Wagner, & Preston, 2013; Martin, McLean, O'Neil, & Köhler, 2013). Nonetheless, the HIP is relatively good at discriminating different allocentric spatial locations (Hassabis, Chu, Rees, Weiskopf, Molyneux & Maguire, 2009), similar memories of movie clips (Chadwick, Hassabis, & Maguire, 2011; Chadwick, Hassabis, Weiskopf, & Maguire, 2010), and the perceived stimulus of two alternatives when a 50% morph is shown in a forced-choice discrimination task (Bonnici et al., 2011). Together, these findings argue against the idea that the HIP detects differences between stimuli merely because it contains more information than other regions. However, additional studies examining representations of features and feature-combinations in a single experiment are required to more directly compare feature and identity information in these regions.

One functional magnetic resonance imaging (fMRI) study of MTL activity during remembering used such an approach (Chadwick et al., 2011). Participants viewed four video clips containing one of two events (a character walking while

---

[1] Note that we use the term *holistic* not to refer to the completeness of an episodic memory (as in Horner, Bisby, Bush, Lin, & Burgess, 2015) but instead to refer to the fact that feature-combinations are not purely a function of individual features.