Routledge
Taylor & Francis Group

Check for updates

COMMENTARY

Modality-specificity is not a necessary condition for grounded semantic cognition: commentary on Calzavarini (2023)

Jamie Reilly[a,b] and Jonathan E. Peelle[c,d,e]

aEleanor M. Saffran Center for Cognitive Neuroscience, Temple University, Philadelphia, PA, USA; bDepartment of Communication Sciences and Disorders, Temple University, Philadelphia, PA, USA; cCenter for Cognitive and Brain Health, Northeastern University, Boston, MA, USA; dDepartment of Communication Sciences and Disorders, Northeastern University, Boston, MA, USA; eDepartment of Psychology, Northeastern University, Boston, MA, USA

I. Introduction

The study of semantic memory has been shaped by numerous disciplines, each with its own idiosyncratic lexicon and methods. Many cognitive neuroscientists today link the origins of semantic memory to the pioneering work of Tulving (1972). However, broader interest in the organisation and neural substrates of human conceptual knowledge can be traced through much of recorded history. Our contemporary understanding of semantic memory rests largely upon a foundation of converging evidence from linguistics, philosophy, and aphasiology. Cognitive neuroscience is a relative newcomer whose primary aim is to identify neurobiological systems supporting cognitive phenomena. One of the major challenges for specifying the neural substrates of semantic cognition has involved retrofitting older cognitive models informed by linguistics and epistemology with a vast amount of data gleaned from modern neuroscience.

Since Tulving (1972), the study of semantic cognition has been punctuated by sudden leaps followed by longer plateaus. Few domains of semantic memory – if any – can be considered definitively "settled" in terms of universal consensus. Moreover, many of debates in this field have been rehashed or recycled under slightly different terminology for decades. Calzavarini (2023) (hereafter Calzavarini) argues that ambiguous nomenclature – specifically, misuse of "modality-specific" – is a rate-limiting factor impeding theory-building in semantic cognition, echoing earlier sentiments about the pitfalls of construct validity and reliability (Dove, 2020; Machery, 2009; Martin, 2016).

One of the cornerstones of the scientific method is falsification (see Popper, 2005). Any viable theory must have the capacity both to reconcile existing data and predict patterns among new sources of empirical sources. Because the study of semantic cognition is inherently multidisciplinary, a crucial step involves calibrating and aligning theoretical and methodological assumptions (see also Poeppel, 2012). Calzavarini's article represents a step towards remedying these challenges. Our understanding of Calzavarini's position is that his main argument rests on a syllogism of the following form:

(a) According to theories of grounded (or embodied) cognition, the meanings of concrete concepts (e.g. table, fish, banana) are represented by features with sensory, motor, and affective salience. These features are distributed either within or proximal to regions of the brain engaged during actual perception or motor execution (e.g. imagined simulation of the redness of an apple engages the same regions of visual cortex active when viewing an apple). Many of these brain regions are thought to reflect modular organisation dedicated to processing one modality (e.g. vision, audition, olfaction). As such, these regions are often referred to as "modality-specific".

(b) An emerging body of empirical evidence supports the claim that many regions of the brain represent information from multiple modalities. As such, modality-specificity in its pure form might not exist.

(c) Since (a) is a logically inconsistent with (b), tenability of grounded cognition should be reassessed in light of recent evidence shifting functional specialisation away from modality-specificity toward local computations that underlie sensorimotor knowledge.

We agree with many of the points advanced by Calzavarini including the necessity for standardising and improving construct validity and reliability. The data the author has presented in support of cross-modal brain organisation are compelling. However, we diverge on the logic that lack of modality-specificity precludes grounded cognition. In the sections to follow we discuss evidence for modality-specificity in the human brain and argue that modality-specificity is not a necessary condition for grounded cognition.

II. Modality-specific reactivation in semantic memory

Theories of semantic memory tend to fall along a spectrum from fully grounded (or embodied) to amodal. A core assumption underlying grounded cognition is that features comprising word and object knowledge are distributed across sensorimotor cortices (Allport, 1985; Cuccio & Gallese, 2018; Gage & Hickok, 2005; Gallese & Lakoff, 2005; Pulvermüller, 2001). In contrast, abstract amodal semantic theories are premised on the claim that object knowledge is mediated by symbols that are ultimately abstracted away from their perceptual roots. Our own perspective, known as the Dynamic Multilevel Reactivation Framework, is a variant of "soft embodiment", incorporating elements of both grounded and amodal semantic theories (Reilly et al., 2016). We hypothesise that regions of the temporal lobe synthesise linguistic with perceptual features transforming object knowledge in a sparse symbolic format. Perceptual simulation (or enactment), a phenomenon we refer to as reactivation, is invoked ad hoc by task demands as needed. For example, given the following questions that rely on semantic knowledge:

(A) Which is a *deeper shape of purple* an eggplant or grape juice?

(B) Which is *sweeter*, an eggplant or grape juice?

one might predict that visual imagery would be engaged to answer Question A, whereas gustatory and olfactory imagery would be invoked to answer Question B.

Language processing is rapid, and many of the words we perceive lack perceptual referents (e.g. consider all the words in the sentence you just read). Therefore, reactivation (or simulation) cannot always be a necessary condition for comprehension. We argue that instead, reactivation is selectively engaged when sensorimotor imagery and semantic working memory are required (e.g. low frequency words and non-canonical events).

The Dynamic Multilevel Reactivation Framework can be considered a hybrid semantic theory in that a core set of hub regions process all concepts, whereas a halo of other cortical regions primarily dedicated to sensorimotor and affective processing act as a supporting cast for enriching sparse representations. In our initial work, we described the set of hub regions as amodal and sensorimotor regions as modality-specific using parallel nomenclature with the Hub-and-Spoke theory of semantic cognition (Chiou & Lambon Ralph, 2019; Patterson & Lambon Ralph, 2016). The question for Calzavarini is whether "modality-specificity" or "unisensory" are indeed accurate descriptors.

One possibility is that the entire brain is a supramodal engine undifferentiated by modality but with local computational specialisation. For example, hypothetical Area A could be reliably activated during functional neuroimaging studies (e.g. MEG, fMRI) when perceiving any spectrally complex sound (e.g. speech, birdsong, music). In addition, damage to Area A might result in apparent dissociations for auditory comprehension. These sources of converging evidence would support the hypothesis that Area A is modality-specific for audition. However, there exists another possible account of the data. The primary function of Area A could also involve processing rapid domain-general spatiotemporal transitions. It would, therefore, be erroneous to characterise Area A as modality specific. We believe that this dichotomy between local computations vs. modality-specificity mirrors a longstanding debate regarding constraints on modularity versus computations in brain and mind (Fodor, 2000).

III. Modality-specificity and modularity in neuroscience

Our understanding is that Calzavarini has equated modality-specificity with strict modularity, the primary assumption of which is that a brain region is functionally specialised for processing one and only one information modality. This dichotomy has historical roots in the distinction between cortical localisation theories versus cortical field theories.

The strongest proponents of cortical localisation theory argue that the brain is massively modular with a functional architecture that is specialised for many discrete cognitive and perceptual functions or modules. Phrenology, for example, represents an extreme application of cortical localisation theory situating highly granular traits (e.g. cautiousness, secretiveness) within specific gyri (Gall, 1835). In contrast, cortical field theories hold that the entire cortex acts as an

undifferentiated functional unit (Bruce, 2001; Flourens, 1824; Lashley, 1929; Tizard, 1959). Perhaps the best-known example of a field theory can be found in Lashley's concepts of equipotentiality (i.e. all regions of the brain contribute equally to all tasks) and mass action (Lashley, 1929). According to mass action, the phenotype or qualitative nature of an impairment secondary to a focal neurological insult cannot reliably be predicted by site of lesion since the entire brain is responsible for executing complex behaviours (for refutations see any work on aphasia including Broca, 1865; Dronkers et al., 2004; Goodglass, 1993; Luria, 1974). According to the mass action principle, only severity (but not qualitative type) of impairment can be predicted by the amount (or mass) of damaged tissue (Lashley, 1958).

Calzavarini presented a compelling body of recent evidence to show that the brain is more cross-modal than explicitly integrated into most grounded models of semantic cognition. In turn, the author questioned whether this evolving body of research justifies an imminent shift away from cognitive modules toward supramodal computations. To be clear, Calzavarini has not advocated for a field theory as extreme as equipotentiality. Rather, the author's apparent position is that modality-specificity might be better characterised an emergent property of specialised computations. This hypothesis finds close parallels in two of the most vigorous recent debates in neuroscience centred around whether the fusiform face area (FFA) and the visual word form area (vWFA) constitute modality-specific regions specialised for perceiving human faces and written words in human ventral temporal cortex.

Computational proponents argued that FFA and vWFA are particularly adept in processing spatiotemporal configurations that are crucial for distinguishing human faces and written words. As such, it would be a mistake to label these regions as selective for faces and words. In contrast, modularity proponents have presented an exhaustive amount of converging data from neuroimaging and neuropsychological dissociations (e.g. alexia secondary to VWFA damage and prosopagnosia secondary to FFA damage) to support claims that these regions constitute functionally specialised modules (Cohen et al., 2000; Dehaene & Cohen, 2011; Kanwisher, 2017; Kanwisher & Yovel, 2006; Kanwisher et al., 1997; McCandliss et al., 2003; Mion et al., 2010; Sergent et al., 1992).

Turning from functionally defined brain regions to neural specialisation, Hubel and Wiesel (1962) also demonstrated modality-specificity in their single unit recording work of the primary visual cortex (V1) of cats. The receptive fields of neurons known as "simple cells" in V1 are tuned to specific line orientations and exhibit spike activity only in the context of this particular stimulus property. In this respect, simple cells in V1 can be considered modality-specific for vision. When a single neuron is selective for one sensory modality, that neuron can be considered unisensory or modality specific. When colonies of neurons are functionally specialised for one domain (e.g. face recognition), such regions have modularity.

If FFA and vWFA constitute modality-specific components of the human visual system, then the brain cannot entirely be considered entirely supramodal. Calzavarini's argument, however, is not whether modality-specificity exists but rather that the human brain is far more cross-modal than often assumed. Many regions previously cast as modality-specific would perhaps more accurately be recast as "modality-preferred". We agree with this assertion. For example, characterising visual association cortex as modality-specific (for vision) would be misleading when the same region also integrates auditory information. The more "meta" issue is whether modality-specificity is in fact a necessary condition grounded cognition.

IV. Modality-specificity is not a necessary condition for grounded cognition

A core assumption underlying embodied (or grounded) cognition is that modality-specific simulations underlie conceptual knowledge. That is, remote recall of objects, actions, and events often requires some degree of re-experiencing, enacting, or reactivating the motor and sensory features that underlie meaning. These perceptuomotor simulations could potentially engage modality-preferred brain regions (e.g. primary motor or visual cortex) or instead engage more distributed association regions along with the multiple demand network. The key distinction is that perceptual simulations are themselves modality-specific but the nature of the functional architecture that supports such simulations (modality-preferred or not) is less relevant. For example, when we ask you to judge whether an eggplant is a deeper shade of purple than grape juice, this question drives a perceptual simulation process that is selectively weighted by colour imagery. Grounded cognition is not predicated upon whether such colour imagery is mediated exclusively by any single brain region. In summary, the question of whether grounded cognition engages primary motor and/or sensory cortex represents a red herring since grounded cognition is not predicated upon such activations.

V. A path forward

The *raison d'etre* of cognitive neuroscience is to specify neural mechanisms for cognitive and perceptual phenomena. Most global models of semantic cognition focus on broad principles (e.g. modality-specificity within networks) with gradual loss of precision and specificity at the microscale levels. A path forward involves eschewing "either/or" framing of modality-specificity in favour of a multisensory perspective. We do not envision that such a shift in nomenclature would be either inconsistent or incompatible with the core assumptions of grounded cognition.

In closing, Calzavarini also discussed the nature and promise of representational semantic spaces. This is currently an area of intense interest across many disciplines (e.g. natural language processing, machine learning). An emerging class of high dimensional semantic spaces is facilitating a jump from low dimensional spaces dominated by sensorimotor features to far more expansive and continuous semantic fields capable of capturing similarity relations between a rich array of affective, interoceptive, and social distinctions (Crutch et al., 2013; Fernandino et al., 2022; Pennington et al., 2014; Reilly et al., 2023) Calzavarini's call to expand representational semantic spaces beyond sensorimotor salience is timely, and perspectives integrating more "abstract" features will likely refine our understanding of semantic memory. High dimensional approaches do not preclude sensorimotor representations or modality-preferential processing.

Disclosure statement

Funding

References

Allport, D. A. (1985). Distributed memory, modular subsystems and dysphasia. In S. K. Newman & R. Epstein (Eds.), *Current perspectives in dysphasia* (pp. 207–244). Churchill Livingstone.

Broca, P. (1865). Location of cerebral functions. Location of articulate language. *Bulletin of the Society of Anthropology (Paris)*, 4, 200–203.

Bruce, D. (2001). Fifty years since Lashley's in search of the Engram: Refutations and conjectures. *Journal of the History of the Neurosciences*, 10(3), 308–318. https://doi.org/10.1076/jhin.10.3.308.9086

Calzavarini, F. (2023). The conceptual format debate and the challenge from (global) supramodality: A position paper. *Language, Cognition, and Neuroscience*.

Chiou, R., & Lambon Ralph, M. A. (2019). Unveiling the dynamic interplay between the hub- and spoke-components of the brain's semantic system and its impact on human behaviour. *NeuroImage*, 199, 114–126. https://doi.org/10.1016/j.neuroimage.2019.05.059

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. a., & Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123(2), 291–307. https://doi.org/10.1093/brain/123.2.291

Crutch, S. J., Troche, J., Reilly, J., & Ridgway, G. R. (2013). Abstract conceptual feature ratings: The role of emotion, magnitude, and other cognitive domains in the organization of abstract conceptual knowledge. *Frontiers in Human Neuroscience*, 7, 186. https://doi.org/10.3389/fnhum.2013.00186

Cuccio, V., & Gallese, V. (2018). A Peircean account of concepts: Grounding abstraction in phylogeny through a comparative neuroscientific perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1752), 20170128. https://doi.org/10.1098/rstb.2017.0128

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, 15(6), 254–262. https://doi.org/10.1016/j.tics.2011.04.003

Dove, G. (2020). More than a scaffold: Language is a neuroenhancement. *Cognitive Neuropsychology*, 37(5-6), 288–311. https://doi.org/10.1080/02643294.2019.1637338

Dronkers, N. F., Wilkins, D., van Valin, R. D., Redfern, B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, 92(1-2), 145–177. https://doi.org/10.1016/j.cognition.2003.11.002

Fernandino, L., Tong, J.-Q., Conant, L. L., Humphries, C. J., & Binder, J. R. (2022). Decoding the information structure underlying the neural representation of concepts. *Proceedings of the National Academy of Sciences*, 119(6), e2108091119. https://doi.org/10.1073/pnas.2108091119

Flourens, P. (1824). *Recherches experimentales sur les propietes et les functions du systeme nerveux*. Crevot.

Fodor, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. MIT Press.

Gage, N., & Hickok, G. (2005). Multiregional cell assemblies, temporal binding and the representation of conceptual knowledge in cortex: A modern theory by a "classical" neurologist, Carl Wernicke. *Cortex*, 41(1998), 823–832. https://doi.org/10.1016/S0010-9452(08)70301-0

Gall, F. J. (1835). *On the functions of the brain and of each of its parts: With observations on the possibility of determining the instincts, propensities, and talents, or the moral and intellectual dispositions of men and animals, by the configuration of the brain and head* (Vol. 1). Marsh, Capen & Lyon.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22(3), 455–479. https://doi.org/10.1080/02643290442000310

Goodglass, H. (1993). *Understanding aphasia* (L. S. Cermak, Ed.). Academic Press, Inc.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual

cortex. *The Journal of Physiology*, *160*(1), 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837

Kanwisher, N. (2017). The quest for the FFA and where it led. *The Journal of Neuroscience*, *37*(5), 1056–1061. https://doi.org/10.1523/JNEUROSCI.1706-16.2016

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2109–2128. https://doi.org/10.1098/rstb.2006.1934

Lashley, K. S. (1929). *Brain mechanisms and intelligence: A quantitative study of injuries to the brain*. https://doi.org/10.1037/10017-000

Lashley, K. S. (1958). Cerebral organization and behavior. *Research Publications of the Association for Research in Nervous & Mental Disease*, *36*, 1–18.

Luria, A. R. (1974). Language and brain: Towards the basic problems of neurolinguistics. *Brain and Language*, *1*(1), 1–14. https://doi.org/10.1016/0093-934X(74)90022-4

Machery, E. (2009). *Doing without concepts*. Oxford University Press.

Martin, A. (2016). GRAPES—Grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. *Psychonomic Bulletin & Review*, *23*(4), 979–990. https://doi.org/10.3758/s13423-015-0842-3

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, *7*(7), 293–299. https://doi.org/10.1016/S1364-6613(03)00134-7

Mion, M., Patterson, K., Acosta-Cabronero, J., Pengas, G., Izquierdo-Garcia, D., Hong, Y. T., Fryer, T. D., Williams, G. B., Hodges, J. R., & Nestor, P. J. (2010). What the left and right anterior fusiform gyri tell us about semantic memory. *Brain*, *133*(11), 3256–3268. https://doi.org/10.1093/brain/awq272

Patterson, K., & Lambon Ralph, M. A. (2016). The hub-and-spoke hypothesis of semantic memory. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (pp. 765–775). Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00061-4

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. https://doi.org/10.3115/v1/D14-1162

Poeppel, D. (2012). The maps problem and the mapping problem: Two challenges for a cognitive neuroscience of speech and language. *Cognitive Neuropsychology*, *29*(1-2), 34–55. https://doi.org/10.1080/02643294.2012.710600

Popper, K. (2005). *The logic of scientific discovery*. Routledge.

Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, *5*(12), 517–524. https://doi.org/10.1016/S1364-6613(00)01803-9

Reilly, J., Litovsky, C., Finley, A. M., & Kenett, Y. N. (2023). Bigram semantic distance as an index of continuous semantic flow in natural language: Theory, tools, and applications. *Journal of Experimental Psychology General*. https://doi.org/10.1037/xge0001389

Reilly, J., Peelle, J. E., Garcia, A., & Crutch, S. J. (2016). Linking somatic and symbolic representation in semantic memory: The dynamic multilevel reactivation framework. *Psychonomic Bulletin & Review*, *23*(4), 1002–1014. https://doi.org/10.3758/s13423-015-0824-5

Sergent, J., Ohta, S., & Macdonald, B. (1992). Functional neuroanatomy of face and object processing: A positron emission tomography study. *Brain*, *115*(1), 15–36. https://doi.org/10.1093/brain/115.1.15

Tizard, B. (1959). Theories of brain localization from Flourens to Lashley. *Medical History*, *3*(2), 132–145. https://doi.org/10.1017/S0025727300024418

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. Academic Press.