Special Issue: Against stored abstractions:
A radical exemplar model of language acquisition

FIRST LANGUAGE

First Language
2020, Vol. 40(5-6) 631–635
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0142723720906233
journals.sagepub.com/home/fla
SAGE

Abstractions are good for brains and machines: A commentary on Ambridge (2020)

Kathryn D. Schuler, Jordan Kodner and Spencer Caplan
University of Pennsylvania, USA

Abstract

In 'Against Stored Abstractions,' Ambridge uses neural and computational evidence to make his case against abstract representations. He argues that storing only exemplars is more parsimonious – why bother with abstraction when exemplar models with on-the-fly calculation can do everything abstracting models can and more – and implies that his view is well supported by neuroscience and computer science. We argue that there is substantial neural, experimental, and computational evidence to the contrary: while both brains and machines can store exemplars, forming categories and storing abstractions is a fundamental part of what they do.

Keywords

Abstraction, child language acquisition, exemplar account, natural language processing, neuroscience

In his article 'Against Stored Abstractions,' Ambridge (2020) uses neural and computational evidence to make his case against abstract representations. His argument is that storing only exemplars is more parsimonious – why bother with abstraction when exemplar models with on-the-fly calculation can do everything abstracting models can and more – and he implies that his view is well supported by neuroscience and computer science. In this commentary we argue that there is substantial neural, experimental, and

Corresponding author:
Kathryn D. Schuler, Department of Linguistics, University of Pennsylvania, Room 314C, 3401-C Walnut Street, Philadelphia, PA 19104, USA.
Email: kschuler@sas.upenn.edu

computational evidence to the contrary: while both brains and machines can store exemplars, forming categories and storing abstractions is a fundamental part of what they do.

Ambridge urges readers not to reject his radical exemplar theory simply because it is at odds with the declarative-procedural model of memory. But the declarative-procedural model is far from the only evidence in favor of abstraction in the brain. To the contrary, a large body of work at all levels of neuroscience – from individual neurons up to neuroimaging data – suggests the brain represents both abstractions and exemplars in a number of ways (e.g., see Seger & Miller, 2010 for a review). For example, neuroscientists have observed that some ensembles of neurons respond to specific exemplars while others respond at the category level. Freedman and colleagues (2003) recorded neurons in two regions – prefrontal cortex and inferior temporal cortex – as monkeys performed a task in which they categorized cats and dogs morphed along a continuum. While the inferior temporal neurons responded selectively to specific stimuli (exemplars), neurons in the prefrontal cortex responded as if they were representing the category: these neurons responded similarly to perceptually distinct exemplars that were members of the same category, but very differently to perceptually similar exemplars that spanned the category boundary.

Follow-up studies have found that, even within a single brain region, neurons are capable of storing both specific exemplars and category-level abstractions. Hippocampal neurons, for example, have been shown to encode not only specific learning instances (O'Reilly & Munakata, 2000) but also category-level information. Hampson and colleagues (2004) trained monkeys to group stimuli into arbitrary categories and found that, rather than 'encoding a mere verbatim representation of individual sensory elements,' hippocampal neurons also responded selectively to category-level features (p. 3184). In the Freedman et al. (2003) work cited above, though the majority of inferior temporal neurons responded to specific stimuli, some of these neurons responded at the category level instead. Because these category-specific inferior temporal neurons responded more slowly than the category-specific prefrontal neurons, some researchers have argued that a feedback loop exists between these two regions, perhaps allowing the prefrontal cortex to feed top-down category-level information back to the inferior temporal cortex (Meyers et al., 2008).

While studies at the level of individual neurons are more commonly done in animals, similar findings have been reported for humans as well. Direct recordings in the human medial temporal lobe have found hippocampal neurons that respond selectively to stimuli at the category level – firing similarly to many perceptually distinct exemplars of the same person, for example (Kreiman et al., 2000). Further, brain imaging studies have found that, just like in monkeys, the human inferior temporal cortex responds more for specific exemplars, while human prefrontal cortex responds more to categories and abstract features (Jiang et al., 2007).

As we have outlined above, one way the brain can implement exemplars plus abstractions is by having different regions or ensembles of neurons respond and represent different levels of structure; another is by having different processing systems operate at different rates. While fast learning from specific exemplars is clearly advantageous, the brain also needs a mechanism to detect the higher-level structures that only emerge across experiences (something we refer to as abstraction). Miller and Buschman (2007) have proposed that the brain accomplishes a balance between fast learning from specific exemplars and slow learning of the structure across them by 'having fast plasticity mechanisms (large changes in synaptic weights) in subcortical structures train slower plasticity (small weight changes) in cortical networks' (Seger & Miller, 2010, p. 210).

So, just as Ambridge urges readers not to dismiss his radical exemplar theory simply because it is not compatible with the declarative-procedural model, we urge you not to dismiss abstractions simply because the brain has a lot of storage space and one fMRI study found an exemplar model explained brain imaging data better than a prototype model (Mack et al., 2013).

We agree that the brain's storage capacity is 'very large' and, in principle, any input information *could* be stored, but we also cannot ignore what evidence from human behavior suggests: we cannot and do not store everything. A tighter bottle-neck to cognitive performance is likely to be found in selective attention to input information rather than in the potential capacity for storage. Human vision has a surprisingly narrow band of foveal focus (see Rayner, 2009 for a review of eye-movements in cognitive processing), and a large literature on the acquisition of categories highlights that learners' real-time gaze is typically limited to only dimensions under active consideration (Rehder & Hoffman, 2005; Shepard et al., 1961). This is in line with learners' *strong* preference to assign category membership to novel stimuli based on limited abstract features rather than holistic 'resemblance' sorting (Medin et al., 1987). Even at the lowest levels of speech generalization, listeners do not store the fine-grained phonetic detail that would be required on a purely exemplar account for any appreciable length of time (Caplan et al., 2019; Jesse & McQueen, 2011). Adaptation to speech variability is limited to abstractions extracted from *first-exposure* to a speaker rather than a statistical or exemplar accumulation of total experience (Kraljic & Samuel, 2007).

If a radical exemplar model is to account for these phenomena, let alone the full gamut of processing results in semantics, syntax, morphology, and phonology, it would need to store representations which are far far richer than even proponents of classic exemplar models normally argue for. The crux of the radical exemplar approach is the existence of an on-the-fly calculation which can account for all phenomena attributed to stored abstractions plus more. However, the number of features that the would-be calculation requires blows up with every new language task.

Even assuming the brain can store all this, it is questionable whether computing across all these features on-the-fly every time an utterance is to be produced is computationally tractable. The examples which Ambridge cites are nothing more than proof of concept, and none has been shown to work at scale. If they did, natural language processing (NLP) may have been solved decades ago – the modern state-of-the-art paradigm for most NLP tasks, deep learning, is essentially souped up connectionism with massive parameter spaces and back propagation. Deep learning models do not aim to accurately represent individual input items and rather achieve their performance by learning complex non-linearly separable abstract classes, for example vector representations of lexical items (Devlin et al., 2018; Mikolov et al., 2013). As Le (2013) shows, individual 'neurons' in deep neural nets respond to abstract visual stimuli such as a cat face, a computational analogy to Freedman et al.'s cat-representing neuron. Information compression, not rich representations of individual training items, is a crucial concept in deep learning, even for models which are designed to remember some aspects of individual inputs (Amodei et al., 2016).

So Ambridge's radical exemplar approach turns out to be less biologically, psychologically, and computationally plausible than he suggests. The premise that exemplars alone, plus on-the-fly calculation, can more parsimoniously account for all the empirical evidence than a model incorporating some kind of stored abstraction becomes tenuous when one concretely works out what kinds of information would have to be included.

ORCID iDs

Kathryn D. Schuler https://orcid.org/0000-0003-2962-731X
Spencer Caplan https://orcid.org/0000-0003-2733-8100

References

Ambridge, B. (2020). Against stored abstractions: A radical exemplar model of language acquisition. *First Language 40*(5-6): 509–559.

Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., Casper, J., Catanzaro, B., Cheng, Q., Chen, G., Chen, J., Chen, J., Chen, Z., Chrzanowski, M., Coates, A., Diamos, G., Ding, K., Du, N., Elsen, E., . . . Zhu, Z. (2016). Deep speech 2: End-to-end speech recognition in English and Mandarin. In N. Lawrence & M. Reid (Eds.). *International Conference on Machine Learning* (pp. 173–182). International Machine Learning Society.

Caplan, S., Hafri, A., & Trueswell, J. (2019, July 24–27). Speech processing does not involve acoustic maintenance. *Paper presented at the 41st Annual Meeting of the Cognitive Science Society*, Montreal, Canada.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1810.04805

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *The Journal of Neuroscience*, *23*(12), 5235–5246.

Hampson, R. E., Pons, T. P., Stanford, T. R., & Deadwyler, S. A. (2004). Categorization in the monkey hippocampus: A possible mechanism for encoding information into memory. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(9), 3184–3189.

Jesse, A., & McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin & Review*, *18*(5), 943–950.

Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., Vanmeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron*, *53*(6), 891–903.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*(1), 1–15.

Kreiman, G., Koch, C., & Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience*, *3*(9), 946–953.

Le, Q. V. (2013, May 26–31). Building high-level features using large scale unsupervised learning. *Paper presented at the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada.

Mack, M. L., Preston, A. R., & Love, B. C. (2013). Decoding the brain's algorithm for categorization from its neural implementation. *Current Biology*, *23*(20), 2023–2027.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology*, *19*(2), 242–279.

Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology*, *100*(3), 1407–1419.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1301.3781

Miller, E. K., & Buschman, T. J. (2007). Rules through recursion: How interactions between the frontal cortex and basal ganglia may build abstract, complex rules from concrete, simple ones. *In Neuroscience of rule-guided behavior* (pp. 419–440). https://doi.org/10.1093/acprof:oso/9780195314274.003.0022

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. The MIT Press.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*(1), 1–41.

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience*, *33*, 203–219.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*(13), 1–42.