Title
Excess Capacity Learning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Dubova, Marina
Sloman, Sabina Johanna

Publication Date
2023

Peer reviewed

Excess Capacity Learning

Marina Dubova (mdubova@iu.edu)
Cognitive Science Program, Indiana University Bloomington
Bloomington, IN 47405 USA

Sabina J. Sloman (sabina.sloman@manchester.ac.uk)
Department of Computer Science, University of Manchester
Manchester, UK M13 9PL

Abstract

Many paradigms in cognitive science posit that human learning is characterized by a limited capacity to represent the information relevant for a given task. We argue that excess capacity — using more representational resources than needed for a task at hand — is a plausible alternative paradigm for the study of human learning. Leveraging recent results from machine learning, we show that excess capacity can be consistent with high predictive ability. We also review extant empirical findings from the cognitive science literature, demonstrating that excess capacity learning can account for a range of empirical phenomena, such as humans' simultaneous yet apparently contradictory tendency to both memorize observations and capture higher-level patterns in them. We conclude by discussing promising directions for future inquiry under the excess capacity learning paradigm.

Keywords: cognitive capacity; overparametrized models; regularization; generalization; compression; simplicity bias

1 Introduction

A body of literature has shown that the amount of information people can deal with at once is constrained by biological and cognitive limitations like working memory capacity, encoding time, and attention (Marois & Ivanoff, 2005; Musslick & Cohen, 2019). Many theories of human cognition posit that the resources people use to represent this information are similarly constrained, i.e., that our internal representations are "simpler" than the information we intake[1]. The goal of this paper is to explore the implications of relaxing the assumption of constrained representational capacity, and taking *excess*, rather than constrained, representational capacity as a conceptual preliminary in the study of human cognition.

To make more concrete the differences between different amounts of representational capacity and introduce the terminology used in the remainder of this paper, consider three different learning systems that can all learn from the same set of observations.

1. The *constrained capacity*[2] system compresses these observations into a lower-dimensional space, effectively discarding details deemed "irrelevant."
2. The *sufficient capacity* system uses as many resources as necessary to reconstruct the observations.
3. The *excess capacity* system expands the observations into a higher-dimensional space, devoting more representational resources than necessary to simply reconstruct the observations: The system's internal representation is of higher dimension, i.e., is "richer," than the data itself.

In cognitive science, we often posit that human learners are constrained capacity systems (Doroudi & Rastegar, 2023). Constrained capacity learning can result in adaptive heuristics (Todd & Gigerenzer, 2012) and reflects bottlenecks induced by limited biological and cognitive resources (Just & Varma, 2007; Miller, 1956; Peterson & Peterson, 1959; Russell & Subramanian, 1994; Waugh & Norman, 1965; Lieder & Griffiths, 2020).

The idea that we apply *excess capacity* — i.e., learn with more resources than the tasks require — is much less explored.[3] Here, we argue that excess capacity learning is worth further exploration as a paradigm for human cognition.

We reassess two types of arguments for constrained cognitive capacity:

1. **Normative** arguments posit that constrained capacity facilitates successful pattern discovery and prediction. In section 2, we show that excess capacity learning can lead learned solutions to make better predictions given new data.
2. **Descriptive** arguments purportedly demonstrate that constrained capacity and the implied simplicity bias is consistent with empirical data on how humans learn. In section 4, we demonstrate that a variety of empirical phenomena in cognitive science can be reinterpreted through the lens of an excess capacity paradigm.

2 Excess capacity learning can lead to high predictive ability

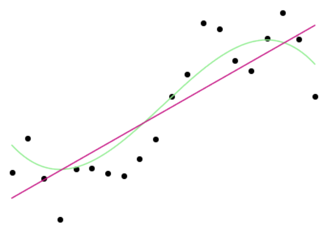
A central challenge of cognitive science is to understand how cognitive systems learn from experience and generalize to sit-

[1] One of the authors has gathered from informal discussions that the focus on constrained capacity learning mechanisms ranges from resource rational analysis (Lieder & Griffiths, 2020) to most applications of drift diffusion (Ratcliff et al., 2016), early connectionist (Rumelhart et al., 1986), symbolic (Raaijmakers & Shiffrin, 1981), and dynamical systems modeling (Thelen & Smith, 1994).

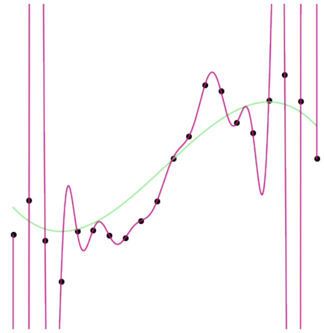
[2] We will use "capacity" to mean "representational capacity" unless otherwise specified.

[3] Neural network models, adapted to tasks like perceptual and language learning, present a prominent recent exception to this (Kriegeskorte, 2015; Yamins & DiCarlo, 2016; Schrimpf et al., 2021). These models often exhibit surprisingly human-like behavior or even outperform humans in these domains.

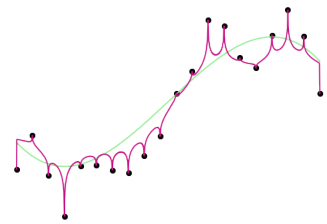
In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.



(a) Degree-one model (constrained capacity).



(b) Degree-twenty model (sufficient capacity).



(c) Degree-one-thousand model (excess capacity).

Figure 1: Polynomial regression models fit to data generated from a degree-three polynomial. Points: Training data. Values on the y-axis are an additive combination of a cubic function of values on the x-axis and Gaussian-distributed noise; in particular, $y = x - .5x^3 + \varepsilon$ where $\varepsilon \sim N(0, .3)$. Green line: Mean of the data-generating function. Violet line: Predictions of the polynomial regression model with degree indicated by the corresponding caption and parameter values fit to the training data. Example is adapted from Nakkiran et al. (2019).

uations they have never encountered before. In both statistical and cognitive learning theory, a hallmark measure for the success of this learning process is *predictive ability*, or the ability to use what one has learned to successfully predict outcomes in unknown situations.[4]

Besides empirical evidence for a simplicity bias in human learning (see section 4), there is a compelling rationale for constrained capacity models on the basis of statistical theory: Until recently, it was widely believed that underparametrized models of the world — models that look at few features or do not transform those features in complicated ways — lead to higher predictive performance (Hastie et al., 2001). The intuition is that, in the face of noise, constrained capacity (underparametrized) models can more reliably capture patterns in the observations. On the other hand, excess capacity (overparametrized) models are prone to overfitting the noise in the observations and thus are commonly thought of as unreliable, in the sense that they can lead to arbitrarily high prediction error.

This is illustrated in Figures 1a and 1b, which show how the predictive ability of a class of polynomial regression models changes as a function of the degree of the corresponding polynomial representation (i.e., the model's representational capacity).[5] Figure 1a shows the predictions of a degree-one polynomial regression model. This model is *constrained* in the sense that it does not have enough parameters to capture the curvature of the true data-generating process. Figure 1b shows the predictions of a degree-twenty polynomial regression model. This model has *sufficient* capacity: It has exactly enough capacity to perfectly reconstruct the training data. The dangers of overfitting can be seen by comparing Figure 1a to Figure 1b. In particular, notice the erratic predictions of the model in Figure 1b in areas it has not seen. Here, there is a trade-off between fitting the noise in the training data, and identifying a pattern that resembles the data-generating process.

Figure 1c shows the predictions of a degree-one-thousand polynomial regression model. This model has *excess* capacity: It has more than enough parameters to perfectly reconstruct the training data. Classic statistical results would lead one to expect this to result in even more erratic predictions than the degree-twenty model. Yet Figure 1c shows exactly the opposite: While the degree-one-thousand model does erroneously fit the noise in the data, it does so in such a way that nevertheless appears to track the underlying structure of the data.[6]

Understanding the mechanisms of the phenomenon that overparameterization increases predictive capacity, known as "benign overfitting," is still an open area of inquiry in machine learning research (Belkin, 2021; Dar et al., 2021). Re-

[4]We use terms like "successful predictions" or "predictive ability" where others might refer to "successful generalization" or "ability to generalize." Some of our arguments will depend on the learner interpolating within the boundaries of the observed data. We avoid the term "generalize" where it might imply extrapolation, or prediction beyond the boundaries of the observed data. While section 7 discusses the importance of extending our arguments to tasks that require extrapolation, others have suggested the ability to interpolate is sufficient for many learning tasks humans encounter in practice (Nastase et al., 2020).

[5]This example was adapted from Nakkiran et al. (2019).

[6]Importantly, note that all three models are fit from the same twenty data points. With reference to discussion in section 1, the learners (here, regression models) are each endowed with the same constraints on the amount of information they can intake, but vary in terms of the resources they apply to represent this information.

sults in linear regression cases like those shown in Figure 1 reveal that overparametrized models can distribute their error in many different directions, which can effectively cancel each other out, resulting in robust behavior of the learned solutions on new observations (Bartlett et al., 2020; Chatterji & Long, 2021). Other results suggest that learning with excess capacity may lead to implicit regularization as the parameters of a model become highly constrained by each other — leading to solutions that are both complex enough to capture all the relevant patterns in the data and less sensitive to noise (Spigler et al., 2019; Geiger et al., 2019).

One objection to the analogy to cognitive systems is that the best-fitting model would be the *correct* model — a degree-three polynomial. Under our definition, this is considered a constrained capacity model. In other words, to recognize what is signal and what is noise, all the learner would have to do is build a good, and relatively simple, internal model that captures the data-generating process and nothing else. However, rarely are we able to know the structure of the true data-generating process — presenting the key challenge of inferring the data-generating process in the wild, which often makes excess capacity models more successful in learning the generalizable patterns in the data than their constrained and sufficient capacity counterparts (Dar et al., 2021; Mallinar et al., 2022; Hastie et al., 2022).

It is also important to note that the phenomenon of benign overfitting does not *always* occur. Excess capacity systems can also overfit in catastrophic ways (Mallinar et al., 2022), resulting in extremely unreliable solutions. Besides elucidating the mechanisms of benign overfitting, the work cited here has also begun to articulate conditions under which benign overfitting is expected to occur. In section 7, we highlight the ways in which this literature can and should play an important role in future work applying the excess capacity paradigm to cognitive science.

3 Properties of learned solutions

As illustrated in Figure 1, the capacity of the learning system affects qualitative properties of learned solutions. Figure 2 is a schematic illustration of how capacity affects three such properties: the expected generalization error of a learned solution, its sensitivity to noise, the degree to which the observed datapoints can be reconstructed (training error), and the complexity of learned solutions (represented by the *x*-axis itself)[7]. Section 2 discussed the double descent exhibited by expected generalization error. The following subsections will discuss the rest of these properties in more detail, emphasizing differences in the properties of solutions recovered by constrained, sufficient and excess capacity systems. In section 4, we discuss existing empirical findings from the study of human cognition in terms of these properties, demonstrating that many findings can be reinterpreted as consistent with the predictions of an excess capacity account.

Memorizing observations while capturing patterns that are robust to noise

Like sufficient capacity systems, and unlike constrained capacity systems, excess capacity systems achieve zero training error, i.e., the learned solutions perfectly reconstruct previous observations.

However, like constrained capacity systems but unlike many sufficient capacity systems, excess capacity systems can simultaneously learn solutions that are robust to noise.[8] With reference to the degree-twenty model in Figure 1b, the sufficient capacity model exhibits high dependence on the particular datapoints that were used to train it: Were we to resample twenty new datapoints on which to train a new degree-twenty model, the new model's predictions would look very different from those shown in Figure 1b. On the other hand, Figures 1a and 1c show that the constrained and excess capacity models exhibit low sensitivity to noise: Models trained on twenty new datapoints drawn from the same distribution would make similar predictions on unseen inputs (Somepalli et al., 2022).

Complex and redundant solutions

Excess capacity systems, by definition, apply more resources to transform observations in different ways. While constrained capacity systems effectively compress the information they intake, excess capacity systems instead expand the information in more varied and sometimes convoluted ways.

4 Rethinking empirical findings through an excess capacity framework

We now briefly review empirical findings from the cognitive science literature, showing that many can be accounted for by an excess capacity framework.

Memorizing observations while capturing patterns that are robust to noise

As discussed in section 3, excess capacity systems perfectly memorize the information they intake, while sometimes also learning a good approximation to general patterns in this information. Similarly, humans have demonstrated a tendency to simultaneously memorize observations and capture higher-level patterns in these observations.

Instance-based effects refer to the observation of our superior ability to recite solutions corresponding to previously-seen items, indicating memorization of individual observations. For example, people's responses to previously-seen items tend to be faster and more accurate (Shepard, 1967;

[7] Note that the excess capacity learning suggests a tradeoff between predictive performance and amount of data. Specifically, more data can actually *hurt* generalization performance, as it moves a learner towards the constrained capacity regime (i.e., moves the interpolation threshold, or sufficient capacity regime, further to the right) (Nakkiran et al., 2021).

[8] As stressed in section 2, excess capacity systems often identify solutions that are robust to noise, but this property is by no means universal or guaranteed (Belkin et al., 2019; Hastie et al., 2022; Dar et al., 2021).

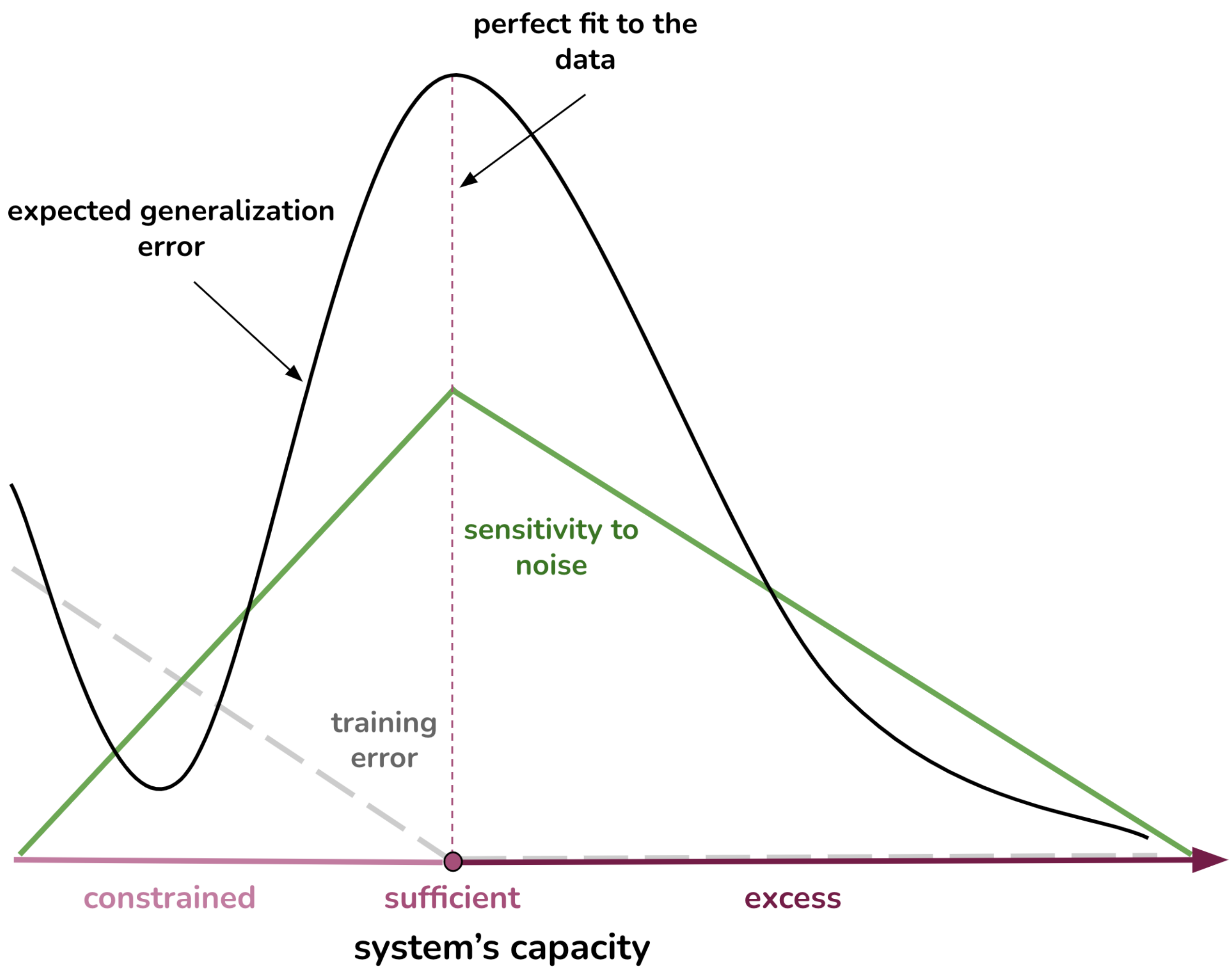


Figure 2: Illustration of the properties of solutions resulting from learning with different levels of cognitive capacity.

Kolers, 1976). Results from a variety of domains indicate that people store nearly all instances from past experience and then use them as reference points when facing novel situations. Instance-based accounts that reflect this have been developed in domains from human categorization (Medin & Schaffer, 1978; Nosofsky, 1988), skill automatization (Logan, 1988), visuomotor learning (Gorman & Goldstone, 2022), dynamic decision-making (Gonzalez et al., 2003; Gonzalez & Dutt, 2011) to problem-solving in groups (Sloman et al., 2021).

Along with storing and using the instances, the evidence also shows that people construct higher-level summary representations (e.g., an average category member) that describe these instances (Murphy, 2004; Schlichting et al., 2015; Bowman et al., 2020). For example, after learning a new category, humans discriminate an average category member which has never been observed faster and categorize it more accurately than any other new instance of the category (Posner & Keele, 1968, 1970), or distort perception towards an average category member (e.g., achromatic bananas are perceived slightly yellow Hansen et al. (2006)) (Kuhl et al., 1992; Goldstone, 1994; Hansen et al., 2006; Bates et al., 2019; Bates & Jacobs, 2020; Dubova & Goldstone, 2021; Harnad, 2003).

Summary-based and instance-based reasoning accounts have been successful at describing learning even within the same domains and tasks, leading to the development of dual-process theories that combine the memorization of instances with a summary or rule inference process (Juslin et al., 2008; Ashby et al., 1998; Erickson & Kruschke, 1998; Briscoe & Feldman, 2011; Hahn & Chater, 1998; Kumaran et al., 2016).

Complex and redundant solutions

Along with being able to capture robust patterns driving the observations, excess capacity learners often take excessively complex ways to get there — their solutions can involve more transformations of information than needed to make accurate predictions. On the other hand, constrained representational capacity implies a preference for simpler explanations or solutions to the problems we face (Chater & Vitányi, 2003). While much work on human learning finds evidence for such a simplicity bias, there is substantial literature suggesting that people rely on redundant cues and exhibit a preference for complex accounts of past observations.

Causal learning People exhibit a large variety of preferences when they assess or induce accounts for the data, sometimes learning what seem to be simple explanations, and sometimes clearly devising complex, and even redundant accounts for the observations. A large body of work has focused on human preferences for simple accounts for the data. For instance, in causal learning with simple stimuli, adults (Lombrozo, 2006, 2007) and preschoolers (Bonawitz & Lombrozo, 2012) prefer simpler causal explanations to more complex ones. When presented with function fits of different complexity, humans prefer simpler functions as generative accounts, often ranking them higher than the medium-complexity ground truth functions (Wilson et al., 2015) (see also Little & Shiffrin (2009) showing a human bias towards linear functions when predicting new points on a graph).

Another body of work, however, demonstrates human preferences for complex and even redundant accounts for the data. For example, when asked to induce a generative rule for items from two artificially created categories, humans are often unable to find the simplest description, inducing overly complex and sometimes redundant rules (Medin et al., 1987). Moreover, when presented with more realistic situations than those used by Lombrozo (2007) and Bonawitz & Lombrozo (2012), adult participants preferred explanations that referred to multiple causes, even if one cause was sufficient to explain the data (Zemla et al., 2017). Other work replicates these results and offers a reconciling "complexity matching hypothesis," positing that people value explanations that match the hypothesized complexity of the domain (Lim & Oppenheimer, 2020). Finally, conspiracy theories present an epitomal example of the human tendency to construct complex and convoluted explanations in an effort to account for every data point (Wojtowicz & DeDeo, 2020).

Perceptual learning Similarly contradictory evidence pervades lower-level implicit learning. It has been argued that perceptual learning proceeds with a simplicity bias — a tendency to infer as little or less structure than needed to accommodate the observations. For instance, when being exposed to stimuli generated by two feature distributions, participants are sometimes more likely to assume that all the stimuli come from just one distribution, especially when the two distributions overlap Gershman & Niv (2013); also, see previously discussed work on categorical perception: Kuhl et al. (1992); Goldstone (1994).

Along with demonstrations of a simplicity bias in perceptual learning, the literature is also full of evidence demonstrating the human tendency to infer more structure than is needed to capture what has been observed. For example, multiple studies show that people posit new categories or clusters to account for small variations in new stimuli — as demonstrated by contrastive perceptual adaptation effects, under which perception is biased away from the stimuli that have already been observed (Leopold et al., 2001, 2005; Dubova & Moskvichev, 2019). Finally, when learning relationships between properties of objects (e.g., gem appearance and price), humans have been shown to often be unable to ignore irrelevant perceptual cues (e.g., gem color), even when they are explicitly told about their irrelevance for the task (Thomson & Oppenheimer, in preparation).

5 When is overfitting benign?

Two of the qualitative properties we highlighted — high predictive ability and the recovery of patterns that are robust to noise — are often exhibited by excess capacity systems (Bartlett et al., 2020; Hastie et al., 2022; Belkin, 2021; Belkin et al., 2019). However, overparameterization can also lead to unstable predictions (Mallinar et al., 2022). This leads to the crucial question of *when* the excess capacity framework leads to high predictive ability — and what the implications are for its use as a framework for cognitive science, given that humans' ability to make good predictions also depends on the task and environment (DeLosh et al., 1997; Kane & Broomell, 2020; Zhu et al., 2009).[9]

Some recent work has begun to elucidate the conditions under which one should expect benign overfitting to occur. For example, there is some evidence that the double descent is more dramatic in instances of model misspecification (Dar et al., 2021) and low noise environments (Hastie et al., 2022). Further exploration of the proposed paradigm should entail developing a better understanding of whether the conditions that predict benign vs. catastrophic overfitting can also predict people's ability to generalize.

6 What about biological limitations?

The presence of biological limitations — the requirement to conserve metabolic and other physiological resources necessary for cognitive processing — has been leveraged as an argument against the excess capacity paradigm (Laughlin, 2001). In an insightful recent opinion piece, Hasson et al. (2020) offer arguments for the plausibility that excess capacity learning might be consistent with the actual representational resources that brains seem to possess. Hasson et al. (2020) argue that it is easier to view the brain as an excess capacity organ learning complex representations that interpolate well, than conceiving of the neural biological limitations as drastically constraining the learning capacity and complexity of learned solutions. While our argument is about representational capacity of cognitive systems in general — which may reflect capacity of not only neural systems, but learning systems broadly considered, which includes embodied minds, machines, and social and human-machine systems (Dubova et al., 2022) — Hasson et al. (2020) present additional evidence that this paradigm is worth exploring.

[9] Some have speculated that human perceptual illusions can also be interpreted as a contextual failure to make good predictions (Kriegeskorte, 2015; Yamins & DiCarlo, 2016; Majaj & Pelli, 2018).

7 Discussion

In this paper, we have argued that excess capacity can be a useful perspective for cognitive scientists to adopt. Here, we discuss some empirical and methodological implications of adopting the excess capacity learning perspective, open directions for future inquiry, and limitations of our argument.

What is left unexplained

Despite being a potentially useful framework for understanding cognition in some situations, excess capacity is not appropriate for capturing the whole body of evidence on how we learn and act. Here, we note just some of the phenomena which are more consistent with limited or sufficient capacity accounts of human cognition.

Along with memorizing instances in some cases, people also routinely forget, with forgetting potentially serving many cognitively adaptive functions (Nørby, 2015). Another body of work looks at how biological systems compress rich analog perceptual signals in their limited capacity sensory organs (Bates & Jacobs, 2020; Stevens, 2013). Here, the evidence suggests that humans and other animals adaptively allocate their sensory and working memory resources to the perceptual dimensions or ranges of stimuli that are most relevant to the tasks they are dealing with (Bates et al., 2019; Goldstone, 1994; Simoncelli & Olshausen, 2001). Similarly, there is evidence that people rely on coarse-grained representations of action and state spaces in dynamic decision-making and planning tasks (Lai & Gershman, 2021) — a presumably essential step for avoiding the curse of dimensionality in such tasks.

Contradicting evidence on limited and excess capacity learning may in part be reconciled by recent work in theoretical neuroscience, which suggests that "compression" and "expansion" might each be at play at different stages of information processing (Farrell et al., 2022; Zhou et al., 2022). Thus, the crucial next step for cognitive science is investigating *when* excess capacity is more reflective of a particular stage of the specific cognitive process at hand, instead of assuming limited capacity for all stages of cognitive processing in all tasks.

What remains to be understood about excess capacity learning?

Excess capacity learning is full of unresolved mysteries. Here, we list just some of the open questions in this field which might be of special interest to cognitive scientists. First of all, as discussed in section 2, the statistical mechanisms driving the ability of overparametrized models to make successful predictions are far from understood. Another promising direction for future empirical and theoretical research is controlled assessment of the ability of excess capacity systems to extrapolate beyond the ranges of observed data. Some work in machine learning, e.g., Dhifallah & Lu (2020); Dar & Baraniuk (2022), suggests that excess capacity hinders learners' ability to extrapolate beyond the boundaries of the data they have observed, but this evidence has so far been limited. Finally, application of the framework to human cognition in dynamic and changing environments would benefit from a better understanding of the learning dynamics of excess capacity systems, looking at how the ability of these systems to find robust patterns in the data changes as learning progresses.

Outlook

Cognitive science benefits from diverse perspectives on cognitive systems (Dale et al., 2022). However, thinking about cognition as having excess, rather than constrained, representational capacity has so far been neglected or even discouraged by the field. Exploring excess capacity learning as a possibility for many domains of cognitive science presents new directions of inquiry for cognitive scientists.

8 Acknowledgments

We thank Sebastian Musslick, Rob Goldstone, Brendan Fleig-Goldstein, Rich Shiffrin, Suyog Chandramouli, YY Ahn, Steven Sloman, Danny Oppenheimer, and three anonymous CogSci reviewers for their comments on the earlier versions of this manuscript.

References

Ashby, F. G., Alfonso-Reese, L. A., Waldron, E. M., et al. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological review*, *105*(3), 442.

Bartlett, P. L., Long, P. M., Lugosi, G., & Tsigler, A. (2020). Benign overfitting in linear regression. *Proceedings of the National Academy of Sciences*, *117*, 30063-30070.

Bates, C. J., & Jacobs, R. A. (2020). Efficient data compression in perception and perceptual memory. *Psychological review*, *127*(5), 891.

Bates, C. J., Lerch, R. A., Sims, C. R., & Jacobs, R. A. (2019). Adaptive allocation of human visual working memory capacity during statistical and categorical learning. *Journal of vision*, *19*(2), 11–11.

Belkin, M. (2021). Fit without fear: remarkable mathematical phenomena of deep learning through the prism of interpolation. *Acta Numerica*, *30*, 203–248.

Belkin, M., Hsu, D., Ma, S., & Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, *116*(32), 15849–15854.

Bonawitz, E. B., & Lombrozo, T. (2012). Occam's rattle: children's use of simplicity and probability to constrain inference. *Developmental psychology*, *48*(4), 1156.

Bowman, C. R., Iwashita, T., & Zeithamova, D. (2020). Tracking prototype and exemplar representations in the brain across learning. *elife*, *9*, e59360.

Briscoe, E., & Feldman, J. (2011). Conceptual complexity and the bias/variance tradeoff. *Cognition*, *118*(1), 2–16.

Chater, N., & Vitányi, P. (2003, Jan). Simplicity: a unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*(1), 19–22. doi: 10.1016/S1364-6613(02)00005-0

Chatterji, N. S., & Long, P. M. (2021). Foolish crowds support benign overfitting. *arXiv preprint arXiv:2110.02914*.

Dale, R., Warlaumont, A. S., & Johnson, K. L. (2022, Nov). The fundamental importance of method to theory. *Nature Reviews Psychology*, 1–12. doi: 10.1038/s44159-022-00120-5

Dar, Y., & Baraniuk, R. G. (2022). Double double descent: On generalization errors in transfer learning between linear regression tasks. *SIAM Journal on Mathematics of Data Science*, *4*(4).

Dar, Y., Muthukumar, V., & Baraniuk, R. G. (2021). A farewell to the bias-variance tradeoff? an overview of the theory of overparameterized machine learning. *arXiv preprint arXiv:2109.02355*.

DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997). Extrapolation: the sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 968.

Dhifallah, O., & Lu, Y. M. (2020). A precise performance analysis of learning with random features. *arXiv preprint arXiv:2008.11904*.

Doroudi, S., & Rastegar, S. A. (2023). The bias–variance tradeoff in cognitive science. *Cognitive Science*, *47*(1), e13241.

Dubova, M., Galesic, M., & Goldstone, R. L. (2022). Cognitive science of augmented intelligence. *Cognitive Science*, *46*(12), e13229.

Dubova, M., & Goldstone, R. L. (2021). The influences of category learning on perceptual reconstructions. *Cognitive Science*, *45*(5), e12981.

Dubova, M., & Moskvichev, A. (2019). Adaptation aftereffects as a result of bayesian categorization. In *Cogsci* (pp. 1669–1675).

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*(2), 107.

Farrell, M., Recanatesi, S., Moore, T., Lajoie, G., & Shea-Brown, E. (2022). Gradient-based learning drives robust representations in recurrent neural networks by balancing compression and expansion. *Nature Machine Intelligence*, *4*(6), 564–573.

Geiger, M., Spigler, S., d'Ascoli, S., Sagun, L., Baity-Jesi, M., Biroli, G., & Wyart, M. (2019). Jamming transition as a paradigm to understand the loss landscape of deep neural networks. *Physical Review E*, *100*(1), 012115.

Gershman, S., & Niv, Y. (2013). Perceptual estimation obeys occam's razor. *Frontiers in Psychology*, *4*.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*(2), 178.

Gonzalez, C., & Dutt, V. (2011). Instance-based learning: integrating sampling and repeated decisions from experience. *Psychological review*, *118*(4), 523.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*(4), 591–635.

Gorman, T. E., & Goldstone, R. L. (2022). An instance-based model account of the benefits of varied practice in visuomotor skill. *Cognitive Psychology*, *137*, 101491.

Hahn, U., & Chater, N. (1998). Similarity and rules: distinct? exhaustive? empirically distinguishable? *Cognition*, *65*(2-3), 197–230.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature neuroscience*, *9*(11), 1367–1368.

Harnad, S. (2003). Categorical perception.

Hasson, U., Nastase, S. A., & Goldstein, A. (2020). Direct fit to nature: an evolutionary perspective on biological and artificial neural networks. *Neuron*, *105*(3), 416–434.

Hastie, T., Montanari, A., Rosset, S., & Tibshirani, R. J. (2022). Surprises in high-dimensional ridgeless least squares interpolation. *The Annals of Statistics*, *50*(2), 949–986.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. Springer.

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition*, *106*(1), 259–298.

Just, M. A., & Varma, S. (2007). The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(3), 153–191.

Kane, P. B., & Broomell, S. B. (2020). Applications of the bias–variance decomposition to human forecasting. *Journal of Mathematical Psychology*, *98*, 102417.

Kolers, P. A. (1976). Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(5), 554–565.

Kriegeskorte, N. (2015). Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annual review of vision science*, *1*, 417–446.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What learning systems do intelligent agents need? complementary learning systems theory updated. *Trends in cognitive sciences*, *20*(7), 512–534.

Lai, L., & Gershman, S. J. (2021). Policy compression: An information bottleneck in action selection. In *Psychology of learning and motivation* (Vol. 74, pp. 195–232). Elsevier.

Laughlin, S. B. (2001). Energy as a constraint on the coding and processing of sensory information. *Current opinion in neurobiology*, *11*(4), 475–480.

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature neuroscience*, *4*(1), 89–94.

Leopold, D. A., Rhodes, G., Müller, K.-M., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1566), 897–904.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and brain sciences*, *43*.

Lim, J. B., & Oppenheimer, D. M. (2020, Apr). Explanatory preferences for complexity matching. *PLOS ONE*, *15*(4), e0230929. doi: 10.1371/journal.pone.0230929

Little, D. R., & Shiffrin, R. (2009). Simplicity bias in the estimation of causal functions. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 31).

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological review*, *95*(4), 492.

Lombrozo, T. (2006, Oct). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464–470. doi: 10.1016/j.tics.2006.08.004

Lombrozo, T. (2007, Nov). Simplicity and probability in causal explanation. *Cognitive Psychology*, *55*(3), 232–257. doi: 10.1016/j.cogpsych.2006.09.006

Majaj, N. J., & Pelli, D. G. (2018). Deep learning—using machine learning to study biological vision. *Journal of vision*, *18*(13), 2–2.

Mallinar, N., Simon, J. B., Abedsoltan, A., Pandit, P., Belkin, M., & Nakkiran, P. (2022). Benign, tempered, or catastrophic: A taxonomy of overfitting. *arXiv preprint arXiv:2207.06569*.

Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in cognitive sciences*, *9*(6), 296–305.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological review*, *85*(3), 207.

Medin, D. L., Wattenmaker, W. D., & Michalski, R. S. (1987). Constraints and preferences in inductive learning: An experimental study of human and machine performance. *Cognitive Science*, *11*(3), 299–339.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

Murphy, G. (2004). *The big book of concepts*. MIT press.

Musslick, S., & Cohen, J. D. (2019). A mechanistic account of constraints on control-dependent processing: Shared representation, conflict and persistence. In *Cogsci* (pp. 849–855).

Nakkiran, P., Kaplun, G., Bansal, Y., Yang, T., Barak, B., & Sutskever, I. (2019). *Deep double descent (cross-posted on openai blog).* `https://windowsontheory.org/2019/12/05/deep-double-descent/`.

Nakkiran, P., Kaplun, G., Bansal, Y., Yang, T., Barak, B., & Sutskever, I. (2021). Deep double descent: Where bigger models and more data hurt. *Journal of Statistical Mechanics: Theory and Experiment*, *2021*(12), 124003.

Nastase, S. A., Goldstein, A., & Hasson, U. (2020, Nov). Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, *222*, 117254. doi: 10.1016/j.neuroimage.2020.117254

Nørby, S. (2015). Why forget? on the adaptive value of memory loss. *Perspectives on Psychological Science*, *10*(5), 551–578.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: learning, memory, and cognition*, *14*(4), 700.

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of experimental psychology*, *58*(3), 193.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of experimental psychology*, *77*(3p1), 353.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental psychology*, *83*(2p1), 304.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological review*, *88*(2), 93.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in cognitive sciences*, *20*(4), 260–281.

Rumelhart, D. E., Hinton, G. E., McClelland, J. L., et al. (1986). A general framework for parallel distributed processing. *Parallel distributed processing: Explorations in the microstructure of cognition*, *1*(45-76), 26.

Russell, S. J., & Subramanian, D. (1994). Provably bounded-optimal agents. *Journal of Artificial Intelligence Research*, *2*, 575–609.

Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature communications*, *6*(1), 8151.

Schrimpf, M., Blank, I. A., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., ... Fedorenko, E. (2021). The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, *118*(45), e2105646118.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*, 156–163.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual review of neuroscience*, *24*(1), 1193–1216.

Sloman, S. J., Goldstone, R. L., & Gonzalez, C. (2021). A social interpolation model of group problem-solving. *Cognitive Science*, *45*(12), e13066.

Somepalli, G., Fowl, L., Bansal, A., Yeh-Chiang, P., Dar, Y., Baraniuk, R., … Goldstein, T. (2022). Can neural nets learn the same model twice? investigating reproducibility and double descent from the decision boundary perspective. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 13699–13708).

Spigler, S., Geiger, M., d'Ascoli, S., Sagun, L., Biroli, G., & Wyart, M. (2019, Nov). A jamming transition from under- to over-parametrization affects generalization in deep learning. *Journal of Physics A: Mathematical and Theoretical*, *52*(47), 474001. doi: 10.1088/1751-8121/ab4c8b

Stevens, M. (2013). *Sensory ecology, behaviour, and evolution*. Oxford University Press.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. MIT press.

Thomson, K., & Oppenheimer, D. (in preparation). Reconsidering miserly behavior: Reducing the number of decision cues does not necessarily conserve cognitive resources.

Todd, P. M., & Gigerenzer, G. E. (2012). *Ecological rationality: Intelligence in the world.* Oxford University Press.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological review*, *72*(2), 89.

Wilson, A. G., Dann, C., Lucas, C., & Xing, E. P. (2015). The human kernel. *Advances in neural information processing systems*, *28*.

Wojtowicz, Z., & DeDeo, S. (2020). From probability to consilience: How explanatory values implement bayesian reasoning. *Trends in Cognitive Sciences*, *24*(12), 981–993.

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, *19*(3), 356–365.

Zemla, J. C., Sloman, S., Bechlivanidis, C., & Lagnado, D. A. (2017). Evaluating everyday explanations. *Psychonomic bulletin & review*, *24*(5), 1488–1500.

Zhou, D., Kim, J. Z., Pines, A. R., Sydnor, V. J., Roalf, D. R., Detre, J. A., … Bassett, D. S. (2022). Compression supports low-dimensional representations of behavior across neural circuits. *bioRxiv*.

Zhu, J., Gibson, B., & Rogers, T. T. (2009). Human rademacher complexity. *Advances in neural information processing systems*, *22*.