Training curricula and structured representations in human and machine learning

Ronald B. Dekker

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Oxford

Trinity 2021

Wolfson College, University of Oxford

Thesis abstract

Humans display a remarkable ability to learn from previous experience. Far from being passively received, however, the training experiences we encounter in modern life are often chosen by ourselves or other people. Existing work shows that some training schedules ("curricula") work better than others in specific situations, but our understanding of *when* and *why* these curricula work is limited, making it difficult to predict curriculum success in novel settings. Compounding this issue, the curricula that subjectively feel most effective are not always optimal for promoting long-term retention or generalization. The main aim of this thesis is to contribute towards the development of computationally-informed theories of curriculum efficacy and mechanism. For this purpose, we make use of behavioral experiments, neural recordings and supervised neural network models.

A key focus of the thesis is on curricula that "start small" by constraining the information that is initially available to the learner. This premise is evaluated in three distinct learning domains. For continuous discriminations such as parametric category structures, initially limiting training to prototypical examples helps. We show that this is at least in part because human internal representations are inherently noisy, and prototypicality acts as a buffer against noise-induced distortions. In tasks that require integrating information from multiple cues, the cue properties emphasized by the training curriculum predict the type of decision rule that human learners use. In a compositional rule learning setting, human learners, but not artificial neural networks, benefit from training on individual rules one at a time. A second aspect of study relates to the fact that in most settings, it is not memorization of training examples, but generalization and transfer to novel problems that are most crucial to adaptive behavior. In this regard, we revive an old idea in cognitive science, that human thought and action are fundamentally compositional. Evaluating models with an explicitly modular structure on the same compositional rule learning task as humans, we establish several correspondences. These include sensitivity to training curriculum, and generalization patterns exhibited by humans, but not by traditional artificial neural networks. Taken together, it is hoped that the work presented in this thesis provides a small step forward in understanding the training schedules that facilitate successful learning.

Table of Contents

Acknowledgements

This work owes much to the involvement of many whose paths crossed mine.

First and foremost, I thank Chris Summerfield. Many scientists are able to inspire minds, but few have Chris's talent for igniting the spark of scientific intrigue in the heart. He has profoundly affected me, and my outlook on science and the world would not be the same without him.

Several hundred days in the lab spent mostly behind a computer screen were made tolerable primarily by the fantastic members of the Summerfield lab. These include Timo, who shares my love for trees, Leonie, who can turn everyday situations into exciting events, Jan, whose humility and kindness inspires me, Fabrice, a unique blend of bluntness and suave, Yinan, the most dedicated person I know, Fabian, my fellow pirate. I cannot name all of you here, but feel fortunate to have learned and laughed with you. I will fondly remember our silly lunch discussions, nightly outings, table tennis and sandwich hunts.

Thank you to Mira and Mirela for the many summertime escapades over the course of the lockdown. You were instrumental in preserving the limited amount of sanity that I possess.

I'm also blessed with parents who have always supported me in pursuing my dreams, no matter how far from home they lead. I am humbled by the example that you have set.

Finally, I thank Lisa, who moved with me across the sea to share all the highs and lows. Everything I do is through her support and love.

Contributions

The thesis begins with a general introduction and ends with a general discussion. These sections are independent work.

The three research chapters (chapters 2-4) form the core of this thesis. My primary supervisor, Prof. Christopher Summerfield, provided input on the conceptualization, design, analysis and writing of each chapter. At various timepoints, the projects described in these chapters have also benefitted from feedback from various members of Prof. Summerfield´s lab and my secondary supervisor, Prof. Gaia Sceriff.

Chapter 2: Juan Balaguer and Timo Flesch provided input on the conceptualization and design. Andrew Saxe provided input on the neural network simulations. Eleanor Holton assisted with collection of the neural data.

Chapter 3: Fabian Otto collected the data for experiments 1, 2, 3, 4 and 7. Mingfang Zhang collected the data for experiment 6.

Chapter 4: Fabian Otto was involved in the design and analysis, and collected all the data.

This work was funded by the Oxford-Wolfson Marriott-Glyn Humphreys Graduate Scholarship in Experimental Psychology and the COVID-19 Scholarship Extensions Fund (CSEF). It has also received indirect funding by the Human Brain Project (HBP) through a grant awarded to Prof. Summerfield.

Chapter 1: General Introduction

When a foal is born, it can stand within two hours, and after 24 it will gallop. On the other hand, before taking their first unsteady steps, a human infant spends their first nine months doing little else than actively taking in the world, absorbing information. Indeed, no other animal spends as long in an underdeveloped state, dependent on their caregivers (Montagu, 1961). Even in adulthood, learning and adaptation remain a core part of our existence, and our knowledge is constantly refined. This singular focus on learning appears to pay off, however, as humans are spectacularly adaptable, and the mature mind is capable of learning in abstract ways believed to be qualitatively different from other species (Penn, Holyoak, & Povinelli, 2008).

While a full account of the aspects of human learning is beyond the scope of any written work, in this thesis I will attempt to shed light on two of its properties. The first of these is how the order and nature of training experiences, together termed a "curriculum", predict learning. In dealing with this subject, care will be taken to focus not just on empirical outcomes, but on general principles and mechanisms subserving curriculum success or failure. The second area of focus is compositional representation. While compositionality is an elusive concept to define, throughout this thesis I will use the term to describe generalization to unseen combinations by leveraging a representation that consists of discrete parts that may be flexibly combined. This may be by representing a single task as consisting of discrete functional parts, or conversely, being able to combine knowledge gained in different settings in some form of integration.

1.1 Building blocks

1.1.1 Category learning

While no two scenes are exactly identical, interpretation becomes tractable by clustering experiences into more abstract groups. For example, with minimal effort we are able to recognize a Maltese and Dobermann as members of the same category, despite their visual dissimilarity. Moreover, the categories we employ are constantly modified and refined. For example, a radiologist may have to

learn to distinguish tumors from scanning artefacts, while a native speaker of Japanese will have to learn the phonetic contrast between the phonemes /r/ and /l/ in order to become proficient in English. As a fundamental element of human cognition, there is a rich literature investigating computational and neural mechanisms that mediate this process. In broad swathes, initial work on this topic concerned itself with identifying a singular computational mechanism that best described human category learning (e.g. Busemeyer, Dewey, & Medin, 1984; Rosch, 1973). Subsequently, several authors have focused on demonstrating that human learning involves not one, but multiple systems of category learning (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Bowman, Iwashita, & Zeithamova, 2020; Medin, Altom, & Murphy, 1984). It has been argued that there is now a new wave of research that assumes that human category learning is multifaceted, and instead focuses on how the different systems interact (Ashby & Maddox, 2011). In the next paragraphs, I will discuss several prominent models of categorization that have emerged in this literature.

Prototype theory and exemplar theory

No discussion of category learning would be complete without mention of two theories that have been pitted against each other in a tug-of-war which has spanned four decades, and remains ongoing (Homa, Sterling, & Trepel, 1981; Hu & Nosofsky, 2021; Nosofsky, 2020).

The first of these is “prototype theory”. Developed in the early ‘70s (Rosch, 1973), its central notion is that all encounters with a category are compounded into a single category average that exemplifies that category. All subsequent category judgements are then based on proximity to these prototypes. For example, a task in which cats and dogs are discriminated would involve representations of a single prototypical cat and dog, and each exemplar would be categorized based on which of these prototypes it is more similar to.

The main alternative that has been proposed is “exemplar theory” (Brooks, Rosch, & Lloyd, 1978; Medin & Schaffer, 1978). In this account, no explicit prototype is constructed. Instead, individual experiences are stored in memory. When a novel stimulus is encountered, relevant exemplars are retrieved and compared to determine the category label of the current stimulus. There is now ample

work both in support and criticism of either theory (meta-analysis favoring exemplar theory: Nosofsky, 1988; meta-analysis on its shortcomings: Smith & Minda, 2000).

On the theoretical level, there are advantages and disadvantages to either method of categorization. Prototype judgments are efficient, as they involve only a single comparison. On the other hand, exemplar theory can more easily account for atypical category members. For example, we can identify bats as mammals and not birds, despite them having wings and being capable of sustained flight. In laboratory settings, the experimenter usually carefully crafts a tractable number of categories to be learned. Applied more broadly, however, one issue is the scalability of prototypes. For example, a grasshopper may be a prototypical "green animal", yet it is not close to the prototypes of either "green" or "animal". Without a proposed productive mechanism, it is unclear how one would maintain a near-infinite set of prototypes for compound concepts (Fodor & Lepore, 1996).

One early argument in favor of exemplar theory was that when the variance of distortions from a prototype is larger during training, high-variance test examples are categorized more accurately (Strange, Keeney, Kessel, & Jenkins, 1970). After a delay, however, performance on shown exemplars decreases at a steeper rate than performance for novel exemplars or the prototype itself (Homa et al., 1981; Posner & Keele, 1970; Strange et al., 1970), suggesting as that an exemplar-based mechanism may operate on shorter time scales, which gives way to a prototype-based mechanism over time. While it has subsequently been shown that an exemplar-based account where memories are corrupted over time can account for these findings (Hintzman & Ludlam, 1980), the general concept that successful category learning involves abstracting commonalities while forgetting differences between exemplars still holds sway over contemporary research (Vlach, 2014)

Comparisons between exemplar and prototype accounts have made somewhat of a resurgence in the 21st century, spurred on by advances in neuroimaging. Indeed, it has been argued that while both theories make many identical behavioral predictions, neural representation may be better suited to adjudicate between models (Mack, Preston, & Love, 2013). Studies have found medial temporal lobe activation in categorization (Nomura et al., 2007) and overlap in the anatomy subserving categorization and recognition (Nosofsky, Little & James, 2012), both taken to implicate exemplar memory recruitment

in categorization. In a different line of research, Bowman and Zeithamova (2018) find that ventromedial anterior hippocampus and ventromedial prefrontal cortex activity is better explained by prototype than by exemplar models, despite the hippocampus's known involvement in episodic memory. A recent paper unifies these perspectives by demonstrating neural correlates of prototype learning during a post-learning test phase, but both exemplar and prototype representations when learners are queried on interleaved test trials during learning (Bowman et al., 2020).

Rule-based and information-integration tasks

A second dominant distinction between category learning mechanisms is the distinction between processes recruited in rule-based and information-integration tasks (Ashby & Maddox, 2005). Rule-based tasks rely on an explicit reasoning process. Often, but not always, these tasks involve a unidimensional decision criterion. In information-integration tasks, information from multiple stimulus dimensions must be combined, and after successful learning, participants generally remain unable to precisely verbalize the correct decision rule. Delaying feedback impairs information-integration based learning, but not rule-based learning. This finding has been interpreted as attention and executive memory being more directly involved in rule-based tasks, making the timing of feedback less important (Maddox, Ashby, & Bohil, 2003). Conversely, reducing the processing time for which feedback is displayed impairs rule-based learning, but not information-integration learning (Maddox, Ashby, Ing, & Pickering, 2004). A similar result is obtained when feedback processing is disrupted by incorporating a concurrent distractor task with high working memory demands (Zeithamova & Maddox, 2006). These behavioral differences suggest that both processes may rely on different neural mechanisms. Evidence on this is conflicted. One early study reports that rule-based learning makes use of medial temporal lobe and prefrontal structures involved in the declarative memory system, whereas information integration learning depends on posterior regions of the caudate (Nomura et al., 2007). However, a subsequent study found that differences were minimal (Milton & Pothos, 2011), and a recent study instead reports greater medial temporal lobe activity in information-integration learning, and that no region was more activated in rule-based learning (Carpenter, Wills, Benattayallah, & Milton, 2016).

1.1.2 Artificial neural networks

Artificial neural networks (ANNs) form a different class of model which has frequently been applied to categorization problems (Krizhevsky, Sutskever, & Hinton, 2012). However, their application is much more general: it has been mathematically proven that ANNs are able to represent any function, given arbitrary width (the number of nodes in a layer) and a non-polynomial activation function (Leshno, Lin, Pinkus, & Schocken, 1993).

History

The field of explicit mathematical modelling of cognition through the interaction of simple elements originated with seminal work from McCulloch & Pitts (1943), who used elementary units whose inputs and outputs were binary, reflecting the all-or-nothing nature of real neuron firing dynamics. One major open question was how appropriate connections between neurons were formed. Here, one influential theory was brought forward by Donald Hebb, who proposed that synaptic plasticity is enacted by strengthening connections between neurons that fire in close temporal proximity (Hebb, 1949), a principle often summarized as “cells that fire together wire together”. These principles would later be combined in Hopfield networks, a model of associative memory (Hopfield, 1982). Furthermore, Hebb’s theory remains greatly influential in contemporary research and theory (Cao, Summerfield, & Saxe, 2020; Detorakis, Bartley, & Neftci, 2019; Golkar, Lipshutz, Bahroun, Sengupta, & Chklovskii, 2020). The first ANN resembling modern architectures was devised by Frank Rosenblatt, who termed his model the ‘perceptron’ (Rosenblatt, 1958). This model would later be expanded upon by introducing a hierarchy of successive layers of neurons (multi-layer perceptrons), overcoming limitations of single-layer perceptrons which were famously described by Minsky & Papert (1969). The approach of employing such general-purpose architectures and learning algorithms to solve cognitive problems has variously been referred to as connectionism (Bechtel & Abrahamsen, 1991; Fodor & Pylyshyn, 1988; Pinker & Prince, 1988), parallel distributed processing (McClelland, Rumelhart, & Group, 1986) and sub-symbolic AI (Willshaw, 1994), although the latter term encompasses other algorithms as well. Another important advance has been backpropagation (Werbos, 1974), a learning algorithm by which the connection strengths between neurons are adjusted proportionally to errors the network makes, and which can be applied efficiently to any number of layers by applying the chain rule of calculus. This

creates a dynamics consisting of a forward pass, in which activity propagates from inputs through the network layers in ascending order, until eventually forming the network's output, and a backwards pass, in which the error between output and ground truth is computed using an appropriate loss function, which is then propagated backwards through all layers and used to update the weights. Of note is that such an error-driven update rule is similar to and inspired by the classical Rescorla-Wagner rule (Rescorla, 1972) used to model learning in classical conditioning experiments, describing associations between stimuli rather than between neurons.

Multi-layer perceptrons

Multi-layer perceptrons (MLPs) remain the canonical form of ANNs to this day, and see widespread use in contemporary research, including the research chapters of this thesis. More specialized architectures may reach higher accuracy levels than MLPs by incorporating inductive biases tailored toward the problem domain. These include convolutional neural networks (for spatial problems) and recurrent neural networks (RNNs, for temporal problems). However, these can be derived from MLPs by modifying the forward pass function (convolutions) or connectivity pattern (RNN). Moreover, even specialized architectures often incorporate a standard MLP in their structure, especially at the readout stage, in which intermediate presentations are transformed into the final outputs (He, Zhang, Ren, & Sun, 2016; Krizhevsky et al., 2012). In general terminology, the neurons of an ANN are referred to as *nodes* or *units*, and their connections as *edges*. Edges typically have weights w_{ij}, where indices i and j indicate the edge's input and output nodes. In addition, each node typically has a bias *b*, setting a base level of activation. Nodes may furthermore be arranged into successive layers (depth), and each layer may contain any number of nodes (width). The intermediate layers are commonly referred to as "hidden layers", as their activity is not directly observable from the inputs and outputs. The norm is that all nodes in each layer project onto all nodes in the next layer, for which reason an MLP layer may also be referred to as a "fully-connected layer". However, some contemporary approaches deviate from this, for example by probabilistically removing a proportion of nodes during training (dropout; Hinton, Srivastava, Krizhevsky, Sutskever, & Salakhutdinov, 2012), or by adding connections between layers more than one layer apart, called skip connections (He et al., 2016). A final ingredient is the activation function θ, which is generally nonlinear and transforms the sum of input activations and bias into

output activation. When a sigmoid nonlinearity is used, this can be likened to the relationship between summated input current and firing rate that exists in biological neural networks. Consequently, the activation of a node *j* in layer *n*+1 may be described as follows:

$$a_{n+1,j} = \theta(b_{n+1,j} + \sum_{i=1}^{width} (a_{n,i} w_{n+1,ij})) \quad \text{(Eq. 1.1)}$$

An illustration of a full MLP is given below in Figure 1-1.

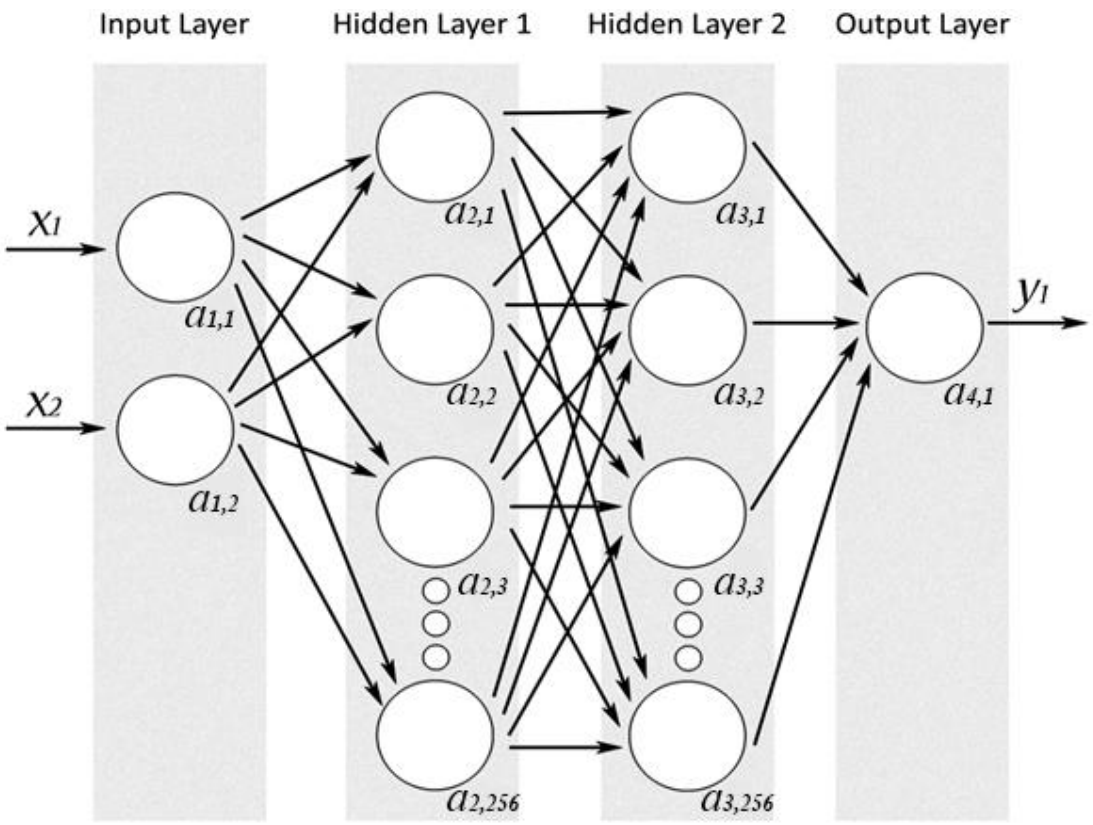

Figure 1-1: Multi-layer perceptron network architecture used in a pilot experiment of research project 1. Inputs are propagated through hidden layer 1, then through hidden layer 2, in order to finally compute output (y). Gray rectangles indicate layers, while circles indicate nodes and arrows indicate edges.

Convolutional neural networks

Recent years have also seen convolutional neural networks (CNNs) become a popular computational model in psychology and neuroscience (Cichy, Khosla, Pantazis, Torralba, & Oliva, 2016; Golan, Raju,

& Kriegeskorte, 2020; Güçlü & van Gerven, 2015; Peterson, Abbott, & Griffiths, 2017). These networks operate by sliding a kernel over 2D input images, dividing them into smaller sections for which the network computes diagnostic features. These features become increasingly complex in deeper layers of the network, yet retain the original spatiotopic structure. Hence, they bear a striking resemblance to visual signal transduction in the mammalian brain, on which they are based (Pinto et al., 2009). In visual cortex, processing relies on receptive fields, which are spatially segregated sections of retinal input. Initially, these are represented as basic features, but representations become increasingly complex along the visual hierarchy while retaining retinotopic structure. It should be noted that CNNs remain vastly different from real brain processing in many respects, including their method of weight updating or "learning", gradient backpropagation, which faces difficulties with biological plausibility (although recent work suggests that a number of traditionally imposed biologically implausible constraints are not crucial to network performance (Akrout, Wilson, Humphreys, Lillicrap, & Tweed, 2019; Lillicrap, Cownden, Tweed, & Akerman, 2016; Whittington & Bogacz, 2019)).

An advantage of these models in modelling psychological tasks is that they can simply be given the stimuli that participants see and receive appropriate feedback to learn. We use convolutional neural networks for this purpose in experimental chapter 1. By contrast, in a model such as an MLP, training directly on real task inputs may be prohibitive, particularly if the stimuli are complex. For example, naturalistic stimuli would likely contain minor shifts in perspective between exemplars, which would form a learning problem of its own MLPs. By contrast, in CNNs translational invariance is built in, as weight maps are shared between all locations. MLPs remain appropriate to model a problem at the cognitive level rather than the perceptual level, as in many tasks, the perceptual aspect is deemed to be non-challenging for real subjects, and the difficulty instead lies in the higher-order problem of how semantic properties map onto responses. In this case, the researcher must specify by hand how the real-world inputs faced by the subject are abstracted as MLP inputs, which leaves many degrees of freedom. To illustrate, consider a canonical task in experimental psychology in which participants are faced with Gabor patches which vary in two dimensions: orientation and spatial frequency. In this case, the researcher may model inputs continuously as two scalars. Alternatively, they may opt for a categorical representation, such as binning dimension values and modelling each dimension as a one-

hot vector, or the experimenter may combine a binned representation with Gaussian inputs, creating inputs that resemble tuning curves. Each of these approaches has their own advantages and disadvantages. Aside from the issue of representational format, the experimenter will also have to select which aspects of the stimulus are represented at all, while the real input may contain additional structure, such as the phase of the Gabor. Convolutional neural networks, by training directly on input images, bypass these issues.

1.1.3 Types of machine learning

Supervised learning

Machine learning approaches are traditionally divided into three types, grouped by the type of feedback signal available to the learner. The first of these is “supervised learning”, in which the correct output is provided after every trial. ANNs generally belong to this category. In such settings, it is standard practice to divide available data into a training set and a test set. The training set is used for iterative minimization of some loss function, for example the squared error between model prediction and ground truth. The most common optimization technique is gradient descent. In this procedure, after the loss is calculated for a single training example (stochastic gradient descent) or a set of training examples (minibatch gradient descent), the first-order derivative of the loss with respect to the parameters of the model is computed. An update is then made in the direction of the negative gradient. Repeating this procedure ensures that a local minimum of the loss function is reached. That is, a constellation of parameter settings where no small change could further decrease the loss. In a class of simple problems termed convex optimization problems, any local minimum is a global minimum. In more complex problems, however, some minima may be better than others, and there is a risk that one may get stuck in a subpar solution. There are several methods to mitigate this risk. Of particular interest for our purposes is the finding that ordering training examples in an appropriate manner can improve both learning speed and the quality of local minima obtained (Bengio, Louradour, Collobert, & Weston, 2009).

For examples in the test set, generalization error is measured, but no gradients are computed. This is done to ensure that the model is not only capable of remembering encountered inputs, but also of

generalizing to novel inputs. However, it should be noted that a core assumption is that data in both sets is independent and identically distributed (i.i.d.). Failure to generalize to out-of-distribution examples is a notorious shortcoming of neural networks (Nagarajan, Andreassen, & Neyshabur, 2020) which we address in experimental chapter 3, where we demonstrate that canonical neural networks succeed at interpolation but not extrapolation problems. This explains why they may do well when trained and evaluated on vast datasets of i.i.d. data, but are prone to error when deviating from the training distribution. Intuitively, one may understand this effect by considering a model which is trained to categorize images of fish and cows. If all fish were in the water, and all cows grazing in a meadow, the model may reach perfect accuracy simply by making use of the background. However, if subsequently an out-of-distribution (OOD) example is queried where a cow has fallen into the water, the network may label that image as containing a "fish". Real-world OOD examples are often more subtle, such as camera quality deteriorating over time or background noise levels varying depending on location. Sensitivity to mild deviations from the training distribution has famously been demonstrated in the form of adversarial examples – adding a minimal amount of curated noise imperceptible to the human eye is sufficient to let a state-of-the art network categorize an image as belonging to a different class (Goodfellow, Shlens, & Szegedy, 2014).

Unsupervised learning

"Unsupervised learning" forms a second group of machine learning approaches in which inputs, but no outputs are provided. Instead, the goal is to find some structure inherent to the inputs. For example, k-means divides the data into a pre-specified number of clusters k by initializing k so-called centroids randomly, and then iteratively assigning all datapoints to their nearest centroid, followed by redefining centroids as the Euclidean mean of all datapoints belonging to that cluster. Such approaches may be used on their own. For example, they can be leveraged for outlier detection (Breunig, Kriegel, Ng, & Sander, 2000), or one could use video subscription service user data to cluster viewers into distinct types for use in movie recommendation. Another general application of unsupervised learning is for estimating latent variables. One class of model leverages supervised techniques by letting a model reproduce its inputs after passing through a sparse bottleneck, necessitating dimensionality reduction. For example, this could be a neural network in which one hidden layer has greatly reduced width, a

setup known as variational auto-encoders. By imposing additional constraints, these models can succeed at extracting disentangled latent factors (Higgins, Matthey, et al., 2017). Incorporating these latents into a supervised learning setting has been shown to facilitate OOD generalization (Higgins, Sonnerat, et al., 2017) and mitigate forgetting when learning multiple tasks in sequence (Flesch, Balaguer, Dekker, Nili, & Summerfield, 2018). Unsupervised methods may also be used more generally to pre-train networks, which has been found to be more effective than starting from randomized weights, a principle known as unsupervised pre-training (Erhan et al., 2010; Krizhevsky, 2009). Intuitively, training a network to represent the principal factors of variation in stimuli may endow it with meaningful features that facilitate learning on the real task.

Reinforcement learning

The third and final main paradigm is “reinforcement learning”, whose main application is solving problems with multiple discrete timesteps, receiving a scalar reward signal as feedback. The observables in each time step are called a state. The model, referred to as an agent, selects from among a pre-defined set of actions in each state, which produce a state-dependent reward and bring about a transition to the next state. The problem or environment is generally a Markov decision process (MDP). This entails that state transitions depend only on the current state and action, and are otherwise independent of all past states and actions. Formally, this is known as satisfying the Markov property. This setup is very flexible and most problems can be cast in a reinforcement learning framework. For example, the canonical supervised classification task described earlier may alternatively be modeled as a reinforcement learning problem in which the input image is the state, output labels are the actions, and transition probabilities are action-independent as they are random draws. However, the main benefit of reinforcement learning lies in its ability to solve multi-step problems with sparse rewards. For example, board games such as chess are amenable to being modeled as MDPs, as the board state is straightforward to characterize and turns naturally discretize time and link transitions to actions. Furthermore, given the present board state, it is irrelevant how it has come to be. This example also exemplifies the issue of sparse rewards. The average chess game ends after around 40 turns, after which you will have either won (positive reward) or lost (negative reward). But to what extent does such an outcome depend on the move made at turn 14 versus that on

turn 36? Two different answers to this credit assignment problem have emerged. One solution is to maintain an explicit model of the transition and reward structure of the environment, allowing the agent to plan ahead. This is "model-based reinforcement learning". An alternative approach that currently sees more use is "model-free reinforcement learning", which foregoes an explicit model of the transition structure. For example, such a model may only maintain the "value" of state-action pairs (TD learning, Sutton & Barto, 2018). While sufficient exploration is important to ensure convergence, once the correct values are learned, control becomes trivial, as one can simply select the actions that have the highest value. The credit assignment problem is solved by assuming that the longer ago an action was taken, the less certain one can be that it contributed to the current outcome. This is commonly implemented by use of a discount factor $0<\gamma<1$, which is exponentiated by the number of steps between the state-action pair that is updated and obtainment of reward. The value, then, is the immediate reward, plus the expected future discounted reward. Generally, model-based reinforcement learning is considered to be computationally costly, but may be more flexible in non-stationary environments. For example, if a single transition probability is changed, such as when one route is closed off in a maze, a new solution can immediately be computed, rather than necessitating re-evaluation of all states. Model-free methods, on the other hand, are generally more efficient and easier to use, but do not afford this flexibility. Among more modern implementations are policy gradient algorithms, a form of model-free reinforcement learning where policies are modeled directly, without necessarily maintaining a value representation (e.g. REINFORCE , Williams, 1992). Hybrid models also exist, for example, in actor-critic algorithms, an actor represents the policy, which learns in a bootstrapped manner from a critic, which estimates values (Konda & Tsitsiklis, 2000).

While even the basic TD algorithm is guaranteed to converge to optimal performance given sufficient training, in practice, sparse rewards and vast state spaces may make training times and compute costs prohibitive. In addition, continuous state and action spaces may require function approximators such as neural networks, which re-introduces the issue of local minima. This suggests that there may be a role for structured training curricula to facilitate learning. We will explore this topic in depth in the section "Curricula in machine learning". One aspect of training unique to the reinforcement learning domain is that agents may be trained either on-policy or off-policy. In the supervised setting, outputs are always

based on the current parameters, allowing the latter to be updated such that the model becomes incrementally more likely to provide the correct response on the queried training examples. This may be considered on-policy learning, as the model that is updated is the same as the one that makes the decisions. By contrast, in reinforcement learning, this need not be the case. Particularly in the model-based setting, given a suitable planning algorithm and sufficient exposure to even a random policy, eventually the true reward and transition structure would be learned, and so the optimal policy can be constructed. A practical limitation is that this method may not be sample-efficient – consider for example the number of games required for a random policy to beat a real chess opponent. Nevertheless, there exist several reasons why one might wish to employ off-policy learning. First, obtaining on-policy data may be costly: a deployed drone can crash, and when training a self-driving car, it would be desirable to minimize accidents and fatalities. Second, interleaving past and current experiences can mitigate forgetting, an approach termed experience replay (Schaul, Quan, Antonoglou, & Silver, 2015). A more subtle point is that one may wish to encourage exploration during training, but not let the mechanisms that encourage exploration distort value estimates. To this end, the well-known Q-learning algorithm updates the value of state-action pairs assuming a greedy policy (deterministically sampling the highest-value action) were followed, regardless of the actual current policy (Watkins & Dayan, 1992). Finally, one may wish to bootstrap learning off an existing model. In a recent example of this, a reinforcement learning agent was trained to superhuman performance on Dota 2, a challenging video game (Berner et al., 2019). As training an agent from scratch was deemed impractical, it was first trained to imitate the behavior of real players.

When the Dota 2 agent from the previous paragraph was trained, each game resulted in a single win or loss, forming a very sparse reward signal. To guide learning, additional rewards were introduced for behaviors that were deemed beneficial, such as defeating enemies and collecting resources. This raises a different question - how should the reward schedule that optimizes learning be determined? While one option is to hand-craft rewards, more structured approaches exist, in which the reward schedule is also learned (Xu, van Hasselt, & Silver, 2018). As in that case, the reward signal is not provided by the environment, it is referred to as an intrinsic reward (Şimşek & Barto, 2006). This has been applied in a meta-learning setting, in which agents have a finite lifetime, during which they learn

an intrinsic reward function, which is then passed to promote learning in the next agent generation (Zheng et al., 2020; Zheng, Oh, & Singh, 2018). These methods are fairly novel, and it seems likely that the upcoming years will bring more clarity as to exactly what is learned in such representations, along with algorithmic improvements. Another interesting method of automated curriculum development has been proposed by Sukhbaatar and colleagues (2017). In their work, they pit agents against each other and observe that as behaviors of either agent become more sophisticated over time, this creates a natural curriculum with ascending complexity. Furthermore, they find that using this self-play as a form of pre-training may also transfer to single-agent target tasks, reducing the number of training episodes required to reach criterion. A similar method has been used in unsupervised learning, where an influential approach has been to train generative models by the interaction of a generator, which produces inputs (e.g. images) and a discriminator, which categorizes whether data were produced by the generator, or were taken from the real dataset (Generative Adversarial Networks; Goodfellow, Pouget-Abadie, et al., 2014). Contemporary versions of this algorithm have been used to produce sophisticated generative models, for example able to produce images of human faces virtually indistinguishable from real humans (Karras et al., 2020).

1.1.4 Classical dichotomies in human learning

Implicit vs. explicit learning

While the above grouping is not without merit when applied to human learners, by tradition, different taxonomies are employed for human and artificial learners. There are many ways to cut up human learning, but none are fully exhaustive and productive. Instead, it is common practice to juxtapose types of learning within specific domains, which differ qualitatively or rely on distinct subsystems. An example of this is this is the aforementioned distinction between exemplar and prototype-based learning in the domain of category learning.

Among the broadest distinctions is that between implicit and explicit learning. While there are different exact definitions of implicit learning (Frensch & Rünger, 2003), it is generally considered to have ties with non-declarative memory and is associated with learning that takes place without conscious awareness (Reber, 1989; Sun, 2008), or that cannot be verbalized (Stadler & Frensch, 1998). In one

famous case study, patient Henry Molaison suffered from anterograde amnesia following hippocampal lobectomy. While he became unable to retain novel explicit memories, he remained able to acquire new motor skills and conditioned responses (Shah, Pattanayak, & Sagar, 2014). Other studies report that subjects respond faster when stimuli obey statistically regular transitions, and may recognize string sequences as being grammatical in artificial grammar learning tasks, both without being able to verbalize their reasoning (Cleeremans, Destrebecqz, & Boyer, 1998).

Explicit learning, on the other hand, is associated with conscious awareness, greater involvement of attention and working memory processes and sometimes with explicit hypotheses or rules (Binder, Hirokawa, & Windhorst, 2009). A further subdivision is made between episodic learning, of personal experiences always bound to a specific time and location, and semantic learning, of more abstract knowledge and facts (Tulving, 1972).

Associative vs. non-associative learning

Another classical distinction is that between associative and non-associative learning. Non-associative learning paradigms focus on changes in response intensity to a single stimulus, generally after repeated exposure. As a result, response intensity may either increase (sensitization) or decrease (habituation) (Fuentes, 2017). Associative learning is a framework applied primarily to animal learning, in which associations are learned between two or more stimuli. This encompasses operant conditioning, a paradigm in which behavioral change is induced through rewards and punishment in the presence of a stimulus, and so may be considered a behavioral psychology analogue of reinforcement learning. In classical conditioning, or Pavlovian conditioning, an unconditioned stimulus which naturally elicits a response, such as fear or salivation, is paired with a neutral stimulus which elicits no response a priori. Through repeated paired exposure, presentation of the neutral stimulus alone becomes sufficient to elicit the response. It is then referred to as a conditioned stimulus, which elicits a conditioned response.

1.1.5 Transfer learning

A world-class guitarist would be able to quickly reach proficiency with the guqin, a Chinese string instrument. This ability to apply what is learned in conditions that differ in some meaningful way from

those encountered during training is referred to as transfer learning. Transfer can be near, such as playing on a different guitar than one is used to, or far and abstract, such as learning to play the piano after having mastered the guitar, or even recognizing that the Copernican system of describing planetary movement through cycles and epicycles resembles a primitive form of Fourier analysis. Transfer learning has been hailed the holy grail of education (Glaser, 1989). Indeed, one may cast doubts on the utility of knowledge that cannot be applied outside the classroom. According to a survey in the office setting, around 38% of employees do not apply what they learned in training immediately post-training, and this figure increases to 66% after a year (Saks, 2002). In a study on analogical reasoning, participants were unable to solve a problem after reading the solution to an analogical problem, unless the similarities were pointed out by the experimenter (Gick & Holyoak, 1980). One perspective known as constructivism addresses such failures of transfer by challenging the assumption that learning is independent of the learner's pre-existing representational schemas and the context in which knowledge and skills are acquired (Prenzel & Mandl, 1993). In line with this, in preschool children, being encouraged to reflect upon or to explain the similarity between problems in their own words promotes transfer to a greater degree than receiving explanations from an experimenter (Brown & Kane, 1988). In addition, it is well-established that human memory is subject to encoding specificity effects whereby retrieval success depends on overlap between ostensibly irrelevant contextual variables in encoding and retrieval (Tulving & Thomson, 1973). While the human capacity for transfer learning is evidently not without limitations and its determinants not fully understood, it remains a source of inspiration for machine learning approaches, where creating learning models capable of human-level transfer learning is an open problem. Illustrating this divide between the level of sophistication in single-task learning versus transfer learning in machines, when a landmark paper demonstrated record performance on a set of seven Atari 2600 videogames (Mnih et al., 2013), this result was obtained through bringing architectural advances in deep learning to the reinforcement learning setting, while the fact that all learned games were structurally related was not capitalized upon. For example, pong and breakout are both games in which the player controls a paddle, and loses when they let the ball pass. However, as the pixel-level patterns are different, forward transfer would have been difficult to obtain, and instead a new model was trained from scratch for each game. More generally, all games involved clearly delineated objects obeying elementary principles, such as

outcomes being determined by collisions between objects, and motion being continuous rather than involving abrupt transitions. One may speculate that a human learner would have picked up on these regularities, and have benefitted from training on one game in learning the next. More generally, transfer situations have been subdivided into transfer of "deep structure", such as different car dashboards looking differently, yet displaying the same information, and "surface structure", such as car and plane dashboards sharing visual features, yet representing different information (Detterman, 1993). Several proposals have suggested that a focus on learning deep structure (sometimes called causal structure or latent structure) is key to obtaining better and more humanlike transfer (Gershman & Niv, 2010; Lake, Ullman, Tenenbaum, & Gershman, 2017). Ironically, it may be said that the deep networks trained by Mnih et al. had succeeded at learning the surface structure of tasks, while their deep structure eluded them.

1.1.6 Definition of compositionality

The term compositionality has been used to refer to disparate phenomena in different fields and by different researchers. Taken literally, all connectionist models can be said to be compositional. The mathematical definition of composition is applying one function to the result of another. That is, given any function f(x) and g(x), their composition (g ∘ f)(x) = g(f(x)). Indeed, in an MLP, $y = \ldots h_2(h_1(x))$, a layer-wise composition, in which the outputs of one layer become inputs to the next. As stated earlier, the sense in which we use the term composition is more constrained, involving representations that consist of discrete parts that may be flexibly combined. However, at least at a low level of description, this property is also inherent to neural networks. Among the most common activation functions are the sigmoid function $(x) = \frac{1}{1+e^{-x}}$, which returns zero for large negative inputs, and the rectified linear unit (ReLu) $\theta(x) = \max(0, x)$, which returns zero for all negative inputs. As a result, for different inputs, different subpopulations of neurons can be "switched off", forming an implicit gating mechanism. Such gating may help in maintaining orthogonal representations for different tasks, and has been brought into connection with real-brain fMRI patterns (Flesch, Juechems, Dumbalska, Saxe, & Summerfield, 2021). Apart from gating, evidence exists for mixed selectivity coding schemes in which distinct properties, such as value and magnitude, may be encoded in orthogonal directions on the

representational manifold. This has been observed in mice (Libby & Buschman, 2021) and macaques (Bernardi et al., 2020), and has also been applied in engineering approaches to sequential task learning (Duncker, Driscoll, Shenoy, Sahani, & Sussillo, 2020; Farajtabar, Azizan, Mott, & Li, 2020). Taken together, compositionality at the level of neurons does not appear well-suited for discriminating between ANN architectures.

At least one other application of the term remains, which is to focus on model behavior – if a model satisfies all desiderata of a 'compositional learner', it may be considered compositional, regardless of the architecture and algorithms involved. In this regard, it has been argued that the human mind exhibits "algebraic compositionality" or "symbolic logic", while neural models are sensitive only to broad patterns over accumulated statistics (Fodor & Pylyshyn, 1988; Lake, Ullman, Tenenbaum & Gershman, 2017; Marcus, 2003). In a recent review, the different possible interpretations of 'compositionality' in terms of model behavior were disentangled into 5 concrete properties, which were evaluated in sequence-to-sequence models. These metrics are (1) systematicity: whether models generalize algebraically on novel combinations of familiar parts, (2) productivity: whether models can generalize to longer input sequences than they were trained on, (3) localism: whether used operations are local or global, (4) substitutivity: whether models treat synonyms identically and (5) overgeneralization: the extent to which models either learn exceptions or overgeneralize rules. Evaluating three popular architectures: a recurrent, a convolution-based and a transformer model, the authors found that all evaluated models exhibited significant limitations on all metrics (Hupkes, Dankers, Mul, & Bruni, 2019). For completion, I will repeat the definition of compositionality maintained throughout this thesis: *generalization to unseen combinations by leveraging a representation that consists of discrete parts that may be flexibly combined*. I do not claim normative superiority over any other definition, and only make use of this definition to specify the scope of the research. Cast into the previously introduced verbiage, the behavioral part of this definition may be framed as satisfying the systematicity principle. In addition, I will focus on mechanisms for selecting between or combining subpopulations or specialized modules. Existing approaches of this kind are detailed in the section "Transfer through structured representations".

1.2 Types of curriculum

1.2.1 Curricula in artificial learning models

In machine learning, the aims of curricula are twofold. First, a well-chosen curriculum can speed up learning. Second, it may act as a continuation method: techniques for obtaining better solutions on non-convex problems (Bengio et al., 2009). While this provides an important starting point, how should appropriate training examples and their optimal order be determined?

The most common approach is to focus on simple training examples first. However, a priori it is unclear how an appropriate and generalizable difficulty score can be determined. One option is to infer ease from loss on the current model (Kumar, Packer, & Koller, 2010). This approach is particularly effective when datasets contain outliers, or high label noise (Meng, Zhao, & Jiang, 2015). Theoretical work proves that convergence of a fresh network on a convex problem is optimized by training on those examples that have the lowest loss for a model that has already learned the task (Weinshall & Amir, 2020). Consequently, there is in theory no need to transition to more difficult examples later on, given sufficient easy example data availability. However, given identical loss on the converged model, learning is fastest when training on those examples that have the highest loss on the current model (Weinshall & Amir, 2020). Taken together, these results suggest that the best training examples are those that are generally easy, but would currently be classified wrongly by the learner. These results have been demonstrated to hold up empirically for non-convex problems and more complex architectures in simulation (Hacohen & Weinshall, 2019).

Other useful difficulty scores may be domain-specific. In reinforcement learning, approaches are being developed to flexibly generate environments (Jain, Isaksen, Holmgård, & Togelius, 2016; Khalifa, Bontrager, Earle, & Togelius, 2020), and parameterized versions of this (Earle, Edwards, Khalifa, Bontrager, & Togelius, 2021) can be used for fine, but domain-specific control over difficulty. In a fairly original experiment, Zaremba and Sutskever (2014) trained RNNs to predict the outcome of arithmetic problems, receiving programming language code as their inputs. These problems varied in the length of the processed integers and the number of code lines to be executed. Comparing several curricula, they found that neither training on the test distribution of full-complexity problems, nor using a

curriculum that gradually increases the complexity of problems was successful. Instead, they obtained the most success when augmenting the easy-to-hard curriculum with interleaved examples of randomized complexity. The authors offer that when training only on simple examples with length *n* integers, the network may learn to use its full capacity to represent the first *n* integers of a number. However, if the length of memoranda later begins to exceed *n*, the entire representational format would have to be overhauled. In assent with this, it had previously been remarked that *"It was obvious in 1971 and even in 1958 that AI programs suffered from a lack of generality. […] The first gross symptom is that a small addition to the idea of a program often involves a complete rewrite beginning with the data structures"* (McCarthy, 1987). Zaremba and Sutskever's finding may be taken as one among several (Chang, Gupta, Levine, & Griffiths, 2018; Hupkes et al., 2019; Lake & Baroni, 2018) illustrations of the more general finding that generalization in neural networks does not satisfy the productivity principle of compositionality.

Somewhat paradoxically, the opposite approach, sometimes called hard example mining (Shrivastava, Gupta, & Girshick, 2016), has also seen success in accelerating learning. Indeed, easier examples generally lead to smaller updates. An extreme example of this is the hinge loss used in soft-margin support vector machines, which is defined to be 0 for correctly classified examples. A method to find the global solution for these models is to train on an iteratively changing subset of examples that are classified incorrectly by the current model (Felzenszwalb, Girshick, McAllester, & Ramanan, 2009). This issue is less severe for more graded loss functions, but it remains the case that examples with lower loss will result in smaller gradients. An alternative solution that has been applied is to selectively increase the loss from easy samples (Qin et al., 2020).

Whether training is facilitated by focusing on easy or difficult examples depends in part on the problem structure. For a parametric structure where dimensions are separated by a hyperplane, focusing on simple examples is effective and sufficient, as this maximizes the angle between the gradient and ground-truth category boundary. Depending on the problem, hard examples may form outliers which are best ignored, or alternatively may form structured classes of exceptions with a learnable structure. An example of the latter would be learning that all whales are mammals. While it is necessary to teach such exception classes to fully refine the decision surface, it seems unlikely that focusing on these

examples early on would place the learner near an optimal basin of attraction or lead to learning generalizable features for classifying all mammals.

A class of approaches related to easy-first curricula employs training schemes where difficulty is held approximately constant. One theoretical analysis suggests that learning is optimized when the training error is kept fixed at around 85% (Wilson, Shenhav, Straccia, & Cohen, 2018). An alternative is to focus on examples that exhibit the greatest variance in classification accuracy between recent minibatches, preventing the learner from training on what it already knows, or on examples that are either too difficult or outliers (Chang, Miller, & McCallum, 2017).

In a final class of curricula, the selection process is agnostic to training example difficulty. When different types of training can be pre-defined in discrete tasks, curricula can be generated by parallel optimization of a teacher which selects tasks, and a learner which trains on these tasks. The teacher can be trained on per-task score change (Matiisen, Oliver, Cohen, & Schulman, 2019), or on additional metrics, such as cross-task accuracy change and model complexity (Graves, Bellemare, Menick, Munos, & Kavukcuoglu, 2017). This approach can be extended to a continuous parametric task space by incorporating Gaussian mixture models (Portelas, Colas, Hofmann, & Oudeyer, 2020).

1.2.2 Training schedules in humans

Massed and spaced practice

Work addressing the question of how the distribution of training episodes over time affects retention in humans originates with Ebbinghaus (1885), who demonstrated that allowing sufficient time between training episodes leads to better retention than massed training, in which the same quantity of training is performed all at once. Several candidate mechanisms for this spacing effect have been proposed. Earlier presentations may induce a transient priming effect, so that in massed presentations, limited processing is required for subsequent recall, leading to a weaker reinforcement of the memory trace (Challis, 1993). Supporting this, the spacing effect is mitigated in paired associate learning when pseudowords are presented in different fonts (Mammarella, Avons, & Russo, 2004), or when faces are presented at different angles (Mammarella, Russo, & Avons, 2002). In second language acquisition, learning words in semantically related clusters abets interference and hinders learning (Tinkham,

1993). An alternative or additional mechanism relies on encoding variability (Cormier, 2014). As contextual variables such as the learner's mental state vary over time, spaced learning ensures that a memory is encoded in a wide variety of contexts, increasing the coverage of contextual variables that may occur at test. Corroborating this view, studying material twice in different locations improves recall relative to studying it twice in the same room (Smith, Glenberg, & Bjork, 1978)

A more nuanced question is how the optimal length of the spacing intervals should be determined. In devising such schemes, the goal is to balance gains in terms of long-term retention associated with longer intervals against the time cost associated with retrieval failures (Pavlik & Anderson, 2008). Modern approximations model the forgetting curve as a power-law function in which the slope decreases the more often an item is trained (Averell & Heathcote, 2011; Kahana & Adler, 2017). In addition, such models commonly assume that different memoranda have different ease factors, by which is meant that some are inherently forgotten at a faster rate than others (Pavlik & Anderson, 2008; Wozniak & Gorzelanczyk, 1994).

Blocked and interleaved practice

Related to the question of how learning experiences should be scheduled over longer timescales is the question how learning experiences are best scheduled within learning episodes. A crucial choice in this regard is whether like items, such as same-category exemplars, or repeated executions of motor behaviors, are clustered into blocks, or interleaved with unlike items. In comparing these schemes, the common outcome is that blocking leads to faster acquisition in terms of training accuracy, but impairs retention (Goode & Magill, 1986; Rohrer & Taylor, 2007; Simon & Bjork, 2001). This effect subverts expectation, as participants aiming for long-term retention reliably estimate blocked training to be the more effective in accomplishing this (Bjork & Bjork, 2011; Kornell & Bjork, 2006). This disconnect between intuition and empirical efficacy is interpreted as subjects using their fluency during training as a proxy for learning efficacy, or mistaking retrieval strength for storage strength (Bjork & Bjork, 1992). The alert reader may have noticed that without further modifications, the interleaving manipulation confounds alternation of categories with longer spacing intervals. As a consequence, retention benefits may naturally fall out of the aforementioned spacing effect. When this is controlled for by increasing the

delay between trials in the blocked condition, a similar retention as in standard blocked presentation is obtained, remaining lower than when an interleaved schedule is used (Taylor & Rohrer, 2010). This suggests that interleaving may promote retention over and above spacing effects.

While retention clearly favors interleaved practice, the picture becomes slightly more complicated when considering which scheme aids learning that permits generalization. Here too, interleaving has been shown to be beneficial, particularly in tasks involving learning a discriminative contrast. These include learning to assign paintings to painters (Kang & Pashler, 2012; Kornell & Bjork, 2006) and distinguishing between bird species (Wahlheim, Dunlosky, & Jacoby, 2011). In other settings, however, learning is found to be improved by blocked training. For example, native speakers of English trained on written French words and their corresponding pronunciation learned better when words were blocked by rule type (Carpenter & Mueller, 2013). Unifying these findings, a theory proposed by Carvalho & Goldstone (2014) posits that human category learning involves comparisons between sequential exemplars. Taking this further, blocking naturally emphasizes commonalities between same-category exemplars, while interleaving emphasizes the differences between categories. Which of these is more helpful for learning ultimately depends on the problem and its underlying structure. Confirming these predictions, Goldstone & Carvalho demonstrated that when same-category exemplars are similar to each other, but distinct from other categories, interleaved training promotes learning. By contrast, if many aspects within categories are different from exemplar to exemplar, such that successful category learning may depend on identifying that which is held constant, participants instead benefit from blocked training. Corroborating that comparisons are the mechanism behind these differences, in a setting where interleaved training outperforms blocked training, simultaneous training on multiple exemplars from different categories leads to a similar performance increase as when exemplars are interleaved, while simultaneous training on same-category exemplars does not induce such a benefit, instead producing accuracy levels comparable to blocked training (Kang & Pashler, 2012). Presumably, the same logic can be applied to the birds from Wahlheim et al.'s study (high within-category similarity, favors interleaving) and to the French word-pronunciation pairs from Carpenter & Mueller, respectively (words vary on many dimensions, so within-category similarity is low, but the systematicity of the rules is highlighted by blocking).

While the primary focus of this section is on trial order effects on learning a single task, a smaller body of work has focused on how interleaving or blocking not categories, but tasks, affects learning. In this domain, one study finds that students trained on mathematical problems perform better when the practice problems interleave problem types. Moreover, after blocked training, students more often mistakenly applied the solution to one type of problem in the wrong context (Taylor & Rohrer, 2010). By contrast, another study involved category learning, where the same stimuli had to be categorized according to orthogonal categorization boundaries in different tasks. Here, subjects performed better at an interleaved test following blocked training than after interleaved training. Moreover, subjects trained on an interleaved curriculum exhibited greater interference, applying the rule from the wrong context more frequently (Flesch et al., 2018). While more work is needed to fully understand these differences, a tentative explanation is that in Flesch et al's study, the appropriate decision boundary in each task can only be learned by integrating information from many trials, which is likely easier in the blocked context. By contrast, in Taylor et al's study, students are pre-trained on understanding the formulas prior to the task, and the problems are self-contained. Instead, the challenge may lie in recognizing when to apply each formula, and the interleaved design may more effectively train to distinguish between these situations.

Active learning

In a number of man-made learning situations, such as university lectures, information is passively received by the learner ('passive learning'; Haidet, Morgan, O'Malley, Moran, & Richards, 2004; Michel, Cater & Varela, 2009). This differs from most naturalistic environments, where learners actively select and seek out their own experiences ('active learning' or 'self-directed learning'). In this section, I will focus on the consequences of allowing learners to choose their own training examples, although I note that the term active learning is sometimes applied more broadly, to include the format of training experiences (such as whether students learn mathematics through lectures or by solving problems; Rosenthal, 1995) and how learners time their study sessions, described in the section "Massed and spaced practice". Intuitively, active learning offers potential advantages, such as access to the most informative examples and facilitating sequential hypothesis testing unfettered by a pre-imposed learning sequence. On the other hand, a potential pitfall of active learning is that the learner may fall

victim to their own cognitive biases, such as the effect described in the previous section whereby learners choose blocked training schedules over interleaved schedules when given the choice, even when this harms their retention.

One influential theory sometimes called ‘optimal experimental design’ takes inspiration from statistical and epistemological accounts of optimal scientific practice to explain a variety of human behaviors, including which questions they ask, how they explore environments, establish causality and test hypotheses (Coenen, Nelson, & Gureckis, 2019; L. Schulz, 2012). Indeed, one study finds that young children engage more with a toy when current information on its function does not disambiguate between different possibilities, suggesting a link between explorative play and devising informative ‘experiments’ (Schulz & Bonawitz, 2007). Whilst this framework provides a useful normative baseline to compare human behavior to and derive predictions from, it has been remarked that its assumptions are somewhat imprecise and not thoroughly investigated, that the framework is not applicable to problems without explicit hypothesis spaces, and that it lacks a clear link to biological and cognitive constraints (Coenen et al., 2019).

To investigate differences between active and passive learning while keeping the objective quality of the examples constant, researchers have used yoked designs. In this approach, one group learns by choosing their own examples, while learners in a second group passively receive the examples chosen by yoked partners in the first group (Gureckis & Markant, 2012). Although it is generally difficult to disentangle the effects active learning may have on motivation from richer effects such as the suitability of the chosen examples for updating the learner’s idiosyncratic beliefs, one study concludes that active learning may help in part by matching chosen examples with the learner’s current hypothesis in a sequential hypothesis testing framework (Markant & Gureckis, 2014).

In closing this section, I should mention that many factors relevant to curriculum success have been omitted, including use of different types of feedback (Ilgen & Moore, 1987; Kepner, 1991; Van Vliet & Wulf, 2006; Vance & Colella, 1990) or different feedback timings (Maddox et al., 2003; Rahmandad, Repenning, & Sterman, 2009), errorless versus error-driven learning (Fillingham, Hodgson, Sage, & Lambon Ralph, 2003; Wilson, Baddeley, Evans, & Shiel, 1994). These topics remain undiscussed here

not because they are less important, but only because they were deemed less pertinent to the work described in this thesis.

1.3 Overarching theories

1.3.1 Starting small

A natural intuition is that training should begin simple, and become more difficult over time (Terrace, 1964). Indeed, when animals are trained on complex tasks, a common procedure is shaping: rewarding simpler and more gradual goals initially, in order to gradually approximate the desired goal (Krueger & Dayan, 2009). The benefits of this procedure are clearest for multi-step problems. For a task consisting of t steps, each with n possible actions, the correct sequence could be any among n^t possibilities. By contrast, by placing the learner in the last step initially and gradually working their way backwards, the solution can be found in a maximum of $n \cdot t$ trials. This means that a task with 10 steps and 10 possible actions per step can be solved in a maximum of 100 trials with a shaping procedure, while without a curriculum, there are 10.000.000.000 possibilities. In tasks without this step-like structure, it is less obvious if and how one may benefit from starting small.

Two difficulties in translating the 'starting small' intuition into a comprehensive theory are that it is not always clear what factors determine difficulty, and that it is not fully understood when training on either hard or difficult examples is most beneficial. One measure of difficulty is the amount of variation itself. In a task where children were trained to throw bean bags a target distance while blindfolded, training on one near and one far target (e.g. 2 and 4 ft) led to higher accuracy on the test distance (e.g. 3 ft) compared to a condition in which the same amount of training was performed on the test distance itself (Kerr & Booth, 1978). Similarly, infants learning minimal word pairs learn the discrimination better when exemplars are generated by different speakers (Rost & McMurray, 2009). By contrast, in a category learning study, a curriculum which limited variation early on promoted learning relative to random sampling (Elio & Anderson, 1984). In the context of category learning, distance to the category boundary can be used as a natural metric of difficulty. Here, training on difficult examples close to the boundary can help for low-dimensional, simple categories (Pashler & Mozer, 2013), information-integration tasks (Spiering & Ashby, 2008) and simple phonetic contrasts (Jongman & Wade, 2007).

However, training on prototypical examples has been found to be more helpful for high-dimensional tasks (Pashler & Mozer, 2013), tasks with unreliable feedback (Patil & Love, 2014) and difficult phonetic contrasts (Jongman & Wade, 2007).

In seminal work, Elman (1993) trained RNNs to process complex sentences, and observed that the networks would fail when trained on full-complexity sentences immediately. By contrast, they succeeded when simpler sentences were introduced first. Alternatively, they were also capable of learning the task when their memory was initially impaired. It was observed that this forms a parallel with work from developmental linguistics, which maintains that maturational constraints are critical in language acquisition (Newport, 1990). While influential, this theory of maturational constraints is not without controversy. On the modeling side, Elman used outdated methods, and the reproducibility of his initial results has been challenged (Rohde & Plaut, 1999). However, in some settings, inducing maturational constraints has been shown to be beneficial even in modern approaches to artificial learning (Karras, Aila, Laine, & Lehtinen, 2017; Morerio, Cavazza, Volpi, Vidal, & Murino, 2017). On the developmental side, there are other factors than maturational constraints that may explain the language learning success of young children (Hakuta, Bialystok, & Wiley, 2003), and adults do not benefit from training under capacity-limiting conditions (Ludden & Gupta, 2000). By comparison, evidence that a training curriculum that starts small can be beneficial is more consistent. In what can be considered a human analogue of Elman's study, humans became more accurate and productive at an artificial grammar learning task when trained first on simple sentences and gradually moving on to complex sentences, compared to when they were trained on the same sentences in a random order (Poletiek et al., 2018).

1.3.2 Transfer through structured representations

In computational models, there exists a balance between model complexity and required data quantity, with more complex models requiring more data to train. What has so far remained elusive to understand and emulate is that humans can perform few-shot learning, learning new concepts and tasks from only one or a few examples, while at the same time being able to leverage rich and complex representations, allowing them to perform a variety of tasks (Lake, Salakhutdinov, & Tenenbaum,

2015). After seeing a single example, humans are capable of classification, communication, imagination and decomposition (Lake et al., 2017). By contrast, state-of-the art classification networks require millions of examples (e.g. ImageNet; 14 million; Deng et al., 2009), yet can usually perform only a single task. Another phenomenon to account for is that human learning is continual. By contrast, when a connectionist network is trained first on one task and then on another, weights for the first task are overwritten, leading to much greater interference than is observed in sequential task learning in humans. This phenomenon is known as catastrophic forgetting (French, 1999; Kirkpatrick et al., 2017). One author observed that a human learner, having mastered a task such as playing the right card in poker, can trivially invert their strategy when instructed to "try to lose". By contrast, in a conventional artificial learner, this process would be cumbersome, requiring gradual adjustment to the inverted strategy over the course of many training examples. Using a structured representation which separates task and the meta-level instruction of trying to lose, such changes can be made fast and reversible, and meta-level instructions can be applied in different domains than where they were first learned (Lampinen & McClelland, 2020).

More generally, what I propose is that a main factor behind these successes is that humans learn across a wealth of experiences in diverse settings which share some structure. The hypothesis is that this leads to task-general representations not as prone to deleterious interference effects, which guide sensible generalization and account for rapid learning and effective knowledge transfer. An open question is what format of structured representation is appropriate. The remainder of this section will focus on describing the state-of-the art in this field.

Among the first computational approaches involving selecting between or combining specialized modules is mixture-of-experts (MoE; Jacobs, Jordan, Nowlan, & Hinton, 1991). In this approach, data is processed in parallel by separate expert networks, followed by a gating mechanism that either adjudicates between (Jacobs et al., 1991) or combines (Jordan & Jacobs, 1992) the outputs of individual experts. In recursive applications of this model, the expert networks themselves may be replaced by MoEs, giving rise to a layered, hierarchical mixture-of-experts (Jordan & Jacobs, 1992). When deep hierarchies are combined with sparse gating mechanisms, models may at the same time have vast representational capacity, while requiring a limited number of computations at deployment, a

principle known as conditional computation (Shazeer et al., 2017). Gating also forms a natural mechanism for preventing catastrophic forgetting, as when different experts are recruited to solve different tasks, there is no gradient for inactive experts, and so no overwriting can take place. By the same token, however, limited overlap in recruited experts may yield fewer opportunities for transfer learning.

Multi-task learning (MTL) is the problem domain of synchronous learning of multiple tasks, and has inspired algorithms with functionally specialized subpopulations. As mentioned previously, there are both costs and benefits to shared task representations. Attempting to leverage the benefits of both coding schemes, several authors have used architectures in which tasks are initially processed jointly in a single network, but then split into parallel, task-specific streams. This method has been used for combining classification and localization (Girshick, 2015), estimating pose and action (Gkioxari, Hariharan, Girshick, & Malik, 2014), face recognition and face landmark detection (Cha Zhang & Zhang, 2014), among other applications. One contribution automates the decision of when networks should bifurcate by introducing 'cross-stitch units', with learnable parameters which select to what extent same-task activations or cross-task representations should be used (Misra, Shrivastava, Gupta, & Hebert, 2016). One limitation of this method is that it requires that a full network be added for every task, limiting scalability.

Aside from training on a variety of objectives that are of direct interest, additional objectives may be introduced solely as a means to better learn the main task. For example, training a convolutional neural network with high-granularity labels, such as apples, oranges and pears, can improve classification performance of low-granularity categories, such as fruit (Chen, Ding, Chin, & Marculescu, 2018). In another example, classification accuracy is increased when a network is trained to perform image reconstruction, rotation prediction and contrastive learning alongside its primary objective (Atito, Awais, & Kittler, 2021). Indeed, a tacit trend in cutting-edge machine learning approaches appears to be to encourage rich representations by incorporating additional self-supervised training objectives. In natural language processing (NLP), the end goal is to perform well on a variety of language understanding tasks (For a common composite benchmark, see GLUE; Wang et al., 2018). Rather than training end-to-end (from inputs to outputs), competitive NLP models are trained on additional

objectives, such as inferring missing words and upcoming sentences (BERT; Devlin, Chang, Lee, & Toutanova, 2018), or probabilistic modeling of unigrams and bigrams (GPT-2; Radford et al., 2019). Similar methods are now beginning to be applied to visual tasks. Facilitating this cross-pollination is the approach of using transformers (Vaswani et al., 2017), originally a tool meant for processing language, in visual tasks. Rather than using sequences of words, these models can also be trained on sequences of neighboring image patches (Dosovitskiy et al., 2020). There is a tradition of learning useful representations from such patches by training models to generate similar representations for different patches taken from the same image (contrastive learning; Becker & Hinton, 1992). In a modern variant of this, a visual transformer was trained to produce similar representations for local image patches as for global patches. Without any supervised training, this model was capable of state-of-the-art object segmentation, its representational space appeared semantically meaningful, and an examination of different parts of the model called attention heads revealed that they were sensitive to different features (Caron et al., 2021). This segmentation of objects into features is reminiscent of the compositional segmentation that humans can perform uninstructed (Biederman, 1987), which is noteworthy as this has previously been used as an example of how human representations, but not supervised network representations, facilitate the performance of a plurality of tasks (Lake et al., 2015).

Another approach which explicitly models recombinable parts are routing networks, an approach which combines 'function blocks', distinct sets of weights at every layer, with a 'router', which selects which function block is applied at every layer of the network. This forms a joint problem of learning functions, and navigation through the appropriate sequence of functions. To train function blocks and the router in tandem, collaborative multi-agent reinforcement learning is used (Rosenbaum, Klinger, & Riemer, 2017). Conceptually, these networks can be viewed as a variant of hierarchical mixture-of-experts where gating takes place before rather than after expert computation. They are also related to PathNets, which use a genetic algorithm rather than reinforcement learning for the routing (Fernando et al., 2017). Recognizing that there are situations which may call for iterative application of the same function, or that the same functions may be applied in different orders in different tasks, a variant of routing networks named Compositional Recursive Learner uses a fixed set of modules rather than layer-specific function blocks, and lets the router apply modules until a maximum recursion depth is

reached, or a special halt module is selected (Chang, Gupta, Levine, & Griffiths, 2018). To make this work, the inputs, outputs and each intermediate step must be in the same format, and domain-specific module types must be pre-specified. We address both of these limitations in chapter 4, under “Meta-learning - recursive modular architecture”.

1.3.3 Link to present work

The starting small principle holds with particular generality for the subset of problems that can be partitioned into successive approximations. For example, in most RL settings, starting close to the goal accelerates learning because information has to be backed up through time. In RNNs, the same principle applies when a delay period is gradually grown. By contrast, in other tasks, such as feedforward supervised learning, it is less obvious why curricula could and should work. Chapter 2 will examine this question in the category learning setting, focusing on the role of distance-to-boundary. This will include fundamental learning dynamics in CNNs, alongside experimentation with human learners. For humans, we focus on the role of noise in internal representations, which may disproportionally affect learning in boundary-proximal settings. Chapter 3 will focus on probabilistic information integration, where information from separate cues must be combined for accurate predictions. Here, curricula that focus on individual cues will be compared with holistic training. In addition, curricula which initially use a limited set of cues are discussed.

The second focus of the thesis is on representations that subserve generalization. This plays some role in Chapter 3, where it will be discussed how different curricula give rise to different integration strategies. The question takes center stage in Chapter 4, which investigates generalization in a compositional rule learning setting. Its behavioral component focuses on how curricula relate to generalization success in human learners, and the type of errors that humans make. These findings inform the two computational components, which focus on representation. The first explores learning is affected by explicitly partitioning it into a chunking problem and a recombination problem of learned chunks. The second computational component is aimed at understanding how rich representations may be acquired through integrating information from diverse tasks.

Chapter 2: Internal noise as a determinant of curriculum efficacy in category learning

2.1 Chapter Abstract

Although the examples used to learn a category can be crafted in many ways and presented in many orders, we currently lack computationally informed theories of why a given curriculum should succeed or fail. The goal of the work presented in this chapter was to better understand the training regimes that facilitate category learning of naturalistic stimuli in healthy human adults.

To this end, we measured human learning on a binary categorization task involving trial-unique, parametrically generated naturalistic tree stimuli. Focusing on the role of distance to the category boundary on learning, across 3 experiments, we found that learning was promoted by boundary-distal training, with robust transfer – subjects trained exclusively on distal examples performed better on boundary-proximal, intermediate and representative distributions of trials than subjects trained directly on these distributions.

Using supervised neural networks as a computational model, we then analyzed how the theoretical learning dynamics are affected by accounting for internal noise occurring at different stages – perception (early noise), decision (late noise) and feedback. Both perceptual noise and decision noise selectively impaired learning in the boundary-proximal regime, qualitatively matching human behavior and thus serving as candidate mechanisms for curriculum effects. Quantitatively, model fitting revealed that human responses were best explained by internal decision noise.

Confirming a causal role for diagnostic examples in promoting learning, per-trial relevant dimension extremity predicted subsequent increases in accuracy. On the neural level, boundary-distal training increased sensitivity to the relevant dimension, but did not affect irrelevant dimension representation. Furthermore, only in the boundary-distal curriculum, the magnitude of post-error FRN (a correlate of

prediction error) predicted the amplitude of subsequent trial CPP (a marker of evidence integration success), providing an empirical demonstration in humans of a recent theory in machine learning which suggests that optimal training examples should lie at the interface of objective simplicity and subjective difficulty.

2.2 Introduction

In this chapter, we focus on categorization, a ubiquitous, well-studied cognitive process that allows us to make sense of a complex world by forming neural representations that are "abstract" or invariant to physical input characteristics, and using them inductively for generalisation or transfer learning (Ashby & Maddox, 2011; Freedman & Assad, 2016). We consider the case where categories are learned by trial-and-error under deterministic feedback, which machine learning researchers call supervised learning.

In humans, the precise nature of category learning task and the sequence of training examples can interact to drive learning in perplexing ways. The notion of starting small – focusing on simpler examples early on in learning – was previously introduced. Intuitively, a curriculum should minimize the time a learner spends on trivially easy or utterly impossibly examples (Pavlik & Anderson, 2008), ensuring they remain in the zone of proximal development (Vygotskiĭ, 1978). Unfortunately, however it is not always a priori obvious what factors make an example "easy" or "hard". Focusing here on the category learning setting, a substantial body of work has investigated the effect of ordering exemplars by similarity (Flesch et al., 2018; Kang & Pashler, 2012; Kornell & Bjork, 2006), variability (Carvalho & Goldstone, 2014; Elio & Anderson, 1984), and distance to boundary (Giguère & Love, 2013; Lee et al., 1988; Pashler & Mozer, 2013; Spiering & Ashby, 2008). However, the results that emerge are contradictory (Carvalho & Goldstone, 2015). For example, when learning to attribute artists to painting styles, interleaving the work of different painters seems to be beneficial (Kang & Pashler, 2012; Kornell & Bjork, 2006), but when classifying gratings (Herzog, Aberg, Frémaux, Gerstner, & Sprekeler, 2012) or naturalistic tree images (Flesch et al., 2018), interleaving of different exemplars or boundaries is detrimental. Similarly, training on more prototypical examples (e.g. those further from the boundary),

which is known as fading, may either help (Giguère & Love, 2013; Pashler & Mozer, 2013) or hurt (Spiering & Ashby, 2008) category learning.

In sum, different training schedules have been proposed to optimize human category learning, whose efficacy may depend on a number of factors, including the use of one-dimensional vs. multidimensional stimuli (Pashler & Mozer, 2013), within- and between-category similarity (Carvalho & Goldstone, 2014), type of categorization task (Spiering & Ashby, 2008) or probabilistic vs. deterministic categories (Giguère & Love, 2013). While empirical results are plentiful, we generally lack computational accounts of why a given curriculum should succeed or fail. To address this gap in knowledge, the goal of the current chapter is to better understand the training regimes that facilitate category learning of naturalistic stimuli in healthy human adults, with a focus on the underlying mechanisms.

In this regard, current theories of curriculum learning have not considered the role of intrinsic variability in neural signal on curriculum effects. Indeed, a vast body of work demonstrates that neural representation features trial-to-trial fluctuations, which are highly correlated between functionally related brain regions (Biswal, Zerrin Yetkin, Haughton, & Hyde, 1995; Cordes et al., 2000; Fox & Raichle, 2007). Moreover, this intertrial variability in representation predicts intertrial variability in behavior across several domains, including motor force in button presses (Fox, Snyder, Vincent, & Raichle, 2007), auditory signal detection (Sadaghiani, Poline, Kleinschmidt, & D'Esposito, 2015) and risk-taking in economic decision making (Chew et al., 2019). Viewed in this context, it appears plausible that decision variables, too, would be subject to intertrial variability. Applied to the category learning setting, such variability might be particularly detrimental when exemplars are close to the category boundary, becoming ambiguous when noise is introduced. Conversely, sufficient distance to the category boundary could confer robustness to noise.

Concretely, we operationalized the question of how exemplar difficulty affects category learning by manipulating distance-to-bound in a high-dimensional stimulus domain (naturalistic images of trees). Behaviorally, we expected participants to learn fastest when exemplars fell far from the boundary, guiding attention to relevant features, and protecting against noise. To disclose a set of minimal constraints that captures this dynamics, we trained a machine classification algorithm in an intentionally

simplified setting, allowing us to explicitly model different types of noise, representing limitations in human learning and decision-making. This allowed us to trace how these noise types interact with curriculum to produce differential effects, as well as the conditions under which their behavior comes to resemble that of humans. On the neural level, we expected boundary-distal training to increase sensitivity to the relevant dimension, reflecting dimension expansion, while we expected sensitivity to the irrelevant dimension to decrease, indicating dimension contraction.

A final hypothesis borrows from recent theoretical work, where it has been proposed that learning is optimized by exemplars that are objectively simple, i.e. have low loss with respect to the optimal hypothesis (such as interpolation training trials), but subjectively difficult, i.e. have high loss with respect to the learner's current hypothesis (Weinshall & Amir, 2020). Fundamentally, under prediction error based learning rules such as the delta rule, updates are in the direction of the input, and their magnitude is proportional to the signed prediction error. Applied to our curricula, as the optimal hypothesis is aligned with the relevant, but not the irrelevant dimension, the average angle between the optimal hypothesis and inputs is lower in interpolation (in which relevant dimension extremity is greater than irrelevant dimension extremity) than in extrapolation (where the converse is true), yielding better generalization. In line with this, we hypothesized that integration success should improve when (a) training on interpolation trials and (b) neural correlates of prediction error are high.

2.3 Methods

2.3.1 Design

In this series of experiments, curriculum was assigned between-subjects. The training curricula were differentiated by their proximity to the category boundary, so that participants received training only on a subset of the relevant parameter space. In all conditions, they were evaluated on test trials which spanned the whole parameter space, but for which no feedback was provided.

Training stimuli were drawn uniformly from intervals of the parameter space bilateral to the category boundary. In the extrapolation condition, only boundary-proximal training examples were used (5-35% of the maximum distance-to-bound in the test set). In interpolation, only boundary-distal training

examples were used (65-95%). In the intermediate condition, an intermediate range was used (35-65%), while in the full uniform control condition, the full range of training examples was used (0-100%). This conditions structure is displayed in the top part of Figure 2-1. We use the term 'interpolation condition' as receiving training only on boundary-distal stimuli means that at test, the learner must generalize to inlying points. By contrast, the boundary-proximal training in the extrapolation condition requires that the learner extrapolate to more distal stimuli at test.

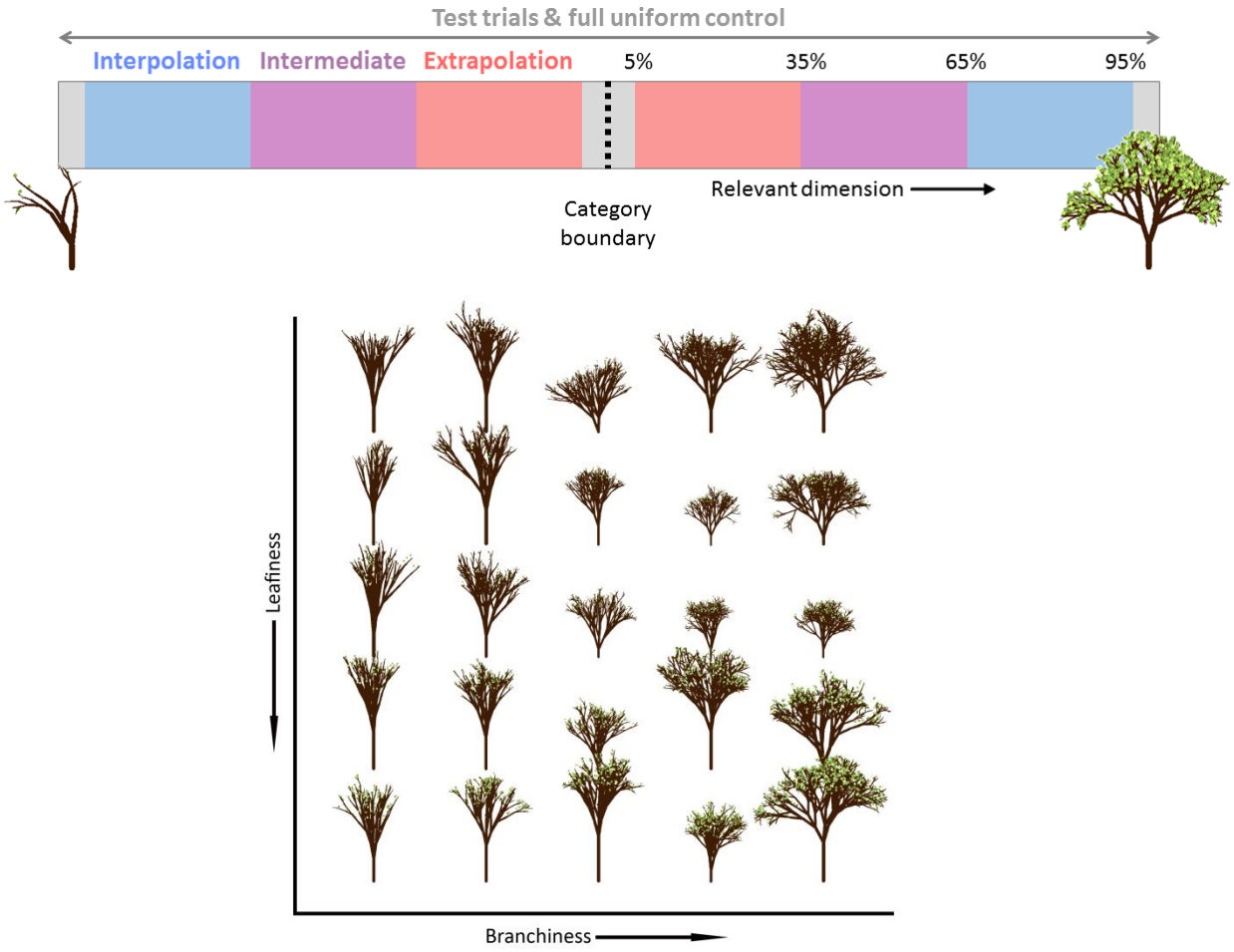

Figure 2-1: Conditions structure and stimulus space. (**Top**) Curriculum structure. Each curriculum uniformly draws samples from a subset of the relevant dimension space. The extrapolation, intermediate and interpolation curricula are non-overlapping and ordered by vicinity to the category boundary. (**Bottom**) Illustration of branchiness and leafiness categories. For each subject, the category boundary was the middle of either the branchiness or the leafiness space.

To chart the temporal dynamics of learning, test trials were interspersed between training trials, without feedback for human participants, and without gradient backpropagation for neural networks. Test trials were sampled uniformly from the entire parameter space in all conditions, as was the irrelevant dimension at both training and test. The generative distributions of stimuli were held constant over time within subjects.

In our main experiment (experiment 1), subjects were tested on-site and assigned to either the interpolation or extrapolation curriculum and EEG was recorded. We confirm and extend the behavioral results in additional online experiments involving a total of 276 subjects. Experiment 2 is an online replication which adds the intermediate condition as a control. Experiment 3 adds a full uniform control condition, and uses non-stationary curricula that gradually move from one extreme to the other over the course of the experiment. As a result, in interpolation-to-extrapolation, the precise same exemplars are shown as in extrapolation-to-interpolation, but in reverse order. Full details on the procedure of experiment 1 and 2 are given here, while experiment 3 is detailed in the supplementary materials.

2.3.2 Stimuli

The stimuli were naturalistic tree images which were procedurally generated using a python-based tree generator. The generator is available at https://github.com/ronald976/RonTree. The stimulus generation procedure is stochastic, such that two stimuli with the same parameters would share relevant visual statistics, yet not be the exact same image. The stimuli are experimentally controlled in two primary dimensions: the number of branches and average angle between branches (together called “branchiness”) and the number of leaves per branch (“leafiness”). Of note is that from the participant´s perspective, these stimuli may feature a larger number of candidate dimensions, such as trunk length, curvature and tree size. The decision criterion was always unidimensional, with the decision boundary lying in the exact middle of that dimension’s axis. In addition to the primary dimensions, stimuli varied randomly in their trunk lengths and were resized to a random percentage between 50-100%. This was done to decouple the branchiness dimension from a simple size estimate, after an online norming pilot found that subjects learned faster in the branchiness than in the leafiness condition without these manipulations. Example stimuli are visualized in Figure 2-1.

2.3.3 Procedure

Condition, relevant dimension (branchiness or leafiness) and key assignment were counterbalanced. The experiment consisted of 14 blocks, each consisting of 20 training trials and 40 test trials. Crucially, training and test trials were interspersed such that there were never more than 2 consecutive training trials, and never more than 7 consecutive test trials. Training trials followed the curriculum structure and were accompanied by feedback, whereas test trials were drawn uniformly from the whole relevant dimension space and were not followed by feedback. The order of test/training trials was yoked between conditions. There was an additional pre-training test phase before the beginning of the experiment, containing 40 test trials.

Stimuli were presented in the center of an LCD monitor. Subjects indicated their choice using the F and J keys. The display was overlaid by a central fixation dot at all times during the experiment. The fixation dot was black by default and turned green, red or gray to indicate positive, negative or no feedback, respectively.

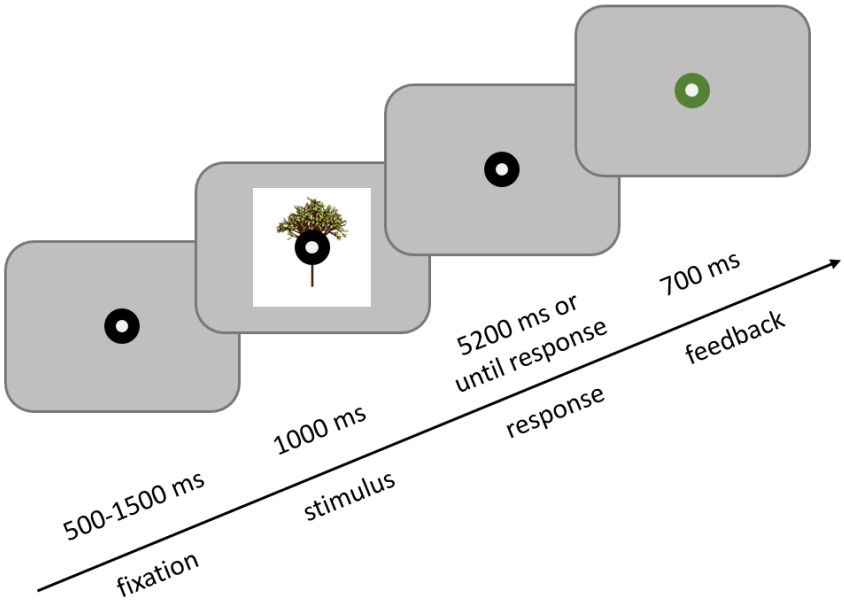

Figure 2-2: Trial structure for experiment 1 and 2.

Trials began with fixation, which lasted for 500ms, plus a 0-1s jitter. Then, the stimulus was displayed for 1s. After stimulus offset, subjects had 5200 ms to respond. After a response or at timeout, a feedback signal was displayed for 700ms. After this, the next trial began. The task design is visualized in Figure 2-2. Between blocks, participants could take a short break, and continue when ready. The design of experiment 2 was identical, but added an intermediate control condition, and subjects were tested online instead of in-lab. The trial structure for experiment 3 was similar, except that stimuli remained on-screen until response, there was no response lockout period, and no jitter between trials.

2.3.4 Participants

In experiment 1, 48 participants (31 female) (mean age 25.8 years, standard deviation 7.25) were recruited through the Oxford Psychology Research (OPR) participant recruitment scheme and local mailing list advertisement. Data were collected in an EEG booth at the University of Oxford. Participants were paid at a rate of £15 per hour, with a minimum of £30.

In experiment 2 and 3, participants were recruited through the online recruitment platform Prolific. An age range of 18-40 years was used. In total, 102 participants (53 female) were recruited for experiment 2, and 174 (91 female) for experiment 3.

All three experiments reported were approved by the Medical Sciences Research Ethics Committee of the University of Oxford

2.3.5 Reporting conventions throughout the thesis

Unless explicitly stated otherwise, all statistics reported throughout this thesis correspond to two-sided tests and compare against a significance level α of .05. Non-parametric tests were used if the data in at least one of the conditions in an analysis violated the assumption of normality. For analyses of neural data, normality was tested using MATLAB's built-in Kolmogorov-Smirnov test. For all other analyses, the normality test from Python's SciPy package was used. The precise outcomes of these normality tests are omitted. Unless otherwise specified, all error bars in figures represent 95% confidence intervals.

Details on model fitting, simulations and EEG acquisition are provided in the supplementary methods.

2.4 Results

Per-subject powerlaw slopes (b parameter in Eq. a.1, supplementary information) were fitted to test accuracy in the first half of the experiment. These were used as a metric of learning rate that is sensitive to trial order. However, we note that all statistical results reported in this section still hold when testing is performed on mean accuracies instead.

In experiment 1, powerlaw slopes were significantly higher in the interpolation group than the extrapolation group, Mann–Whitney U = 652, n_1 = n_2 = 24, p < 0.001. In addition, slopes remained higher in interpolation than in extrapolation when fit only to the most difficult 5-35% of test trials, U = 316, n_1 = n_2 = 24, p = 0.002, which were trials on which the interpolation group was never trained, while extrapolation was trained on them exclusively. The same benefit of interpolation over extrapolation on difficult trials replicated in experiment 2, U = 155, n_1 = n_2 = 34, p < 0.001 and experiment 3, U = 1363, n_1 = 69, n_2 = 72, p < 0.001. In addition, in experiment 2, first-half power-law slopes were higher in interpolation than in intermediate when fitted only to intermediate test trials, U = 331, n_1 = n_2 = 34, p = 0.001. In experiment 3, power-law slopes fitted to all test trials were higher in interpolation than in the full uniform control condition (U = 696, n_1 = 72, n_2 = 33, p < 0.001). The main behavioral results are visualized in Figure 2-3.

2.4.1 Abstract task representation simulations

To gain more insight into the dynamics driving these results, we conducted simulations in an intentionally simplified setting, using multi-layer perceptrons as a computational model. Abstracted inputs were used over real images so that manipulations reflecting the effects of internal noise could be more effectively applied (for an analysis of models which receive the tree stimulus images that participants saw as their inputs, see section a.5). Inputs were abstracted as bivariate Gaussians in a two-dimensional space (50x50). The two dimensions represent a decision-relevant and an orthogonal axis, matching branchiness and leafiness in the original stimuli. The means of the Gaussians were determined by the stimulus parameters. These 50x50 inputs were then flattened to length 2500 vectors and used as inputs for the model. The representation of stimuli as distributions rather than single points in this space can be taken to indicate a belief distribution with varying levels of certainty. This format

has analogies with real-brain neural tuning and activity (Butts & Goldman, 2006), and follows the tradition of probabilistic population codes (Beck et al., 2008). However, we note that whether neural signals are more appropriately modeled through probabilistic population codes or sampling-based methods (Fiser, Berkes, Orbán, & Lengyel, 2010) remains a matter of active debate.

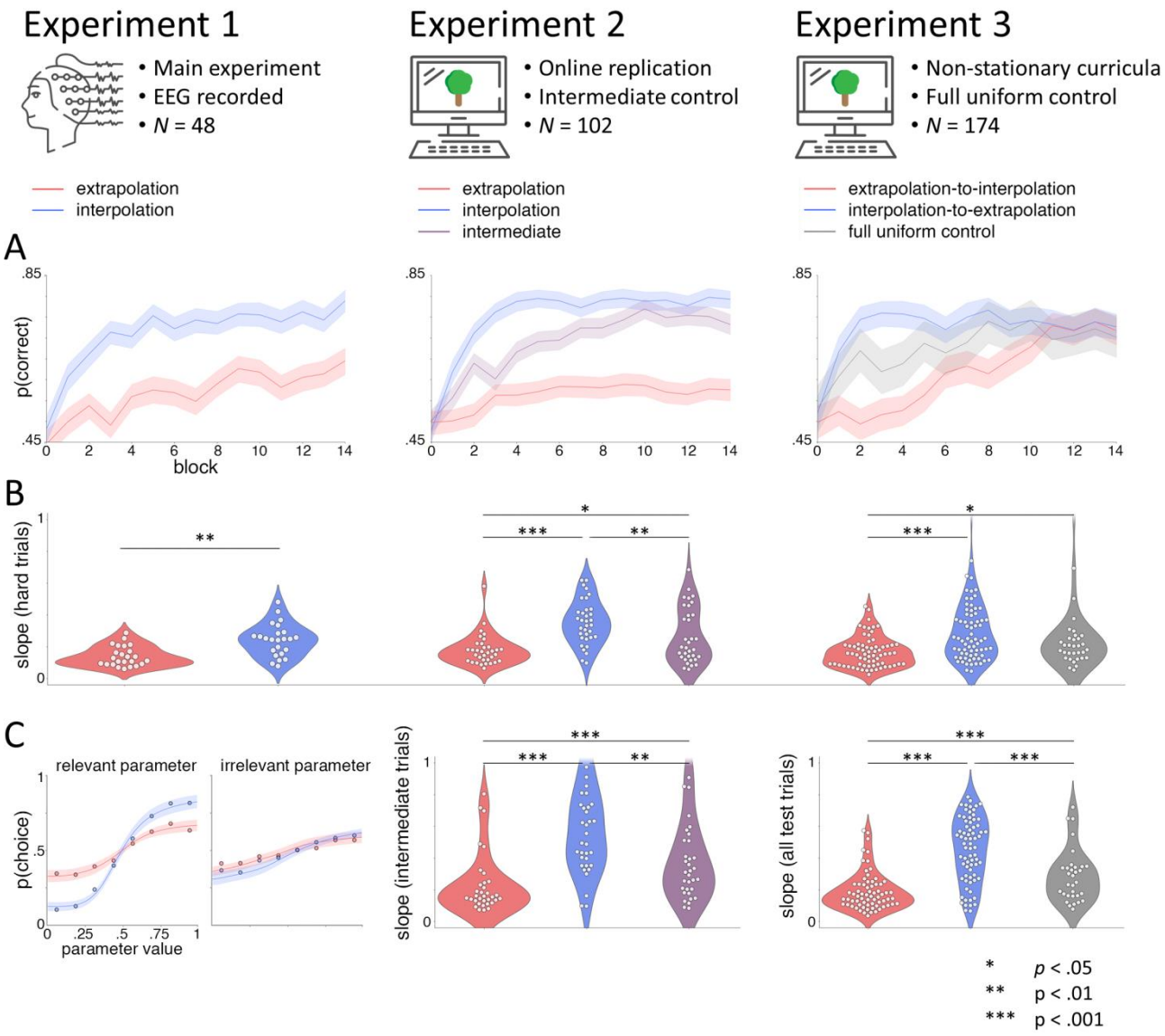

Figure 2-3: Behavioral results. (**A**) Learning curves displaying experimental block number against mean test accuracy. (**B**) Per-subject power-law slopes of hard test trial (5-35% of maximum distance-to-boundary) accuracy in the first half of the experiment as a function of number of training trials completed. Violins represent a kernel density estimation of the underlying distribution. In all conditions, interpolation slopes are significantly higher than extrapolation slopes. (**C**) *Experiment 1*: psychometric curves, displaying choice probability as a function of relevant parameter (left) and irrelevant

parameter (right). The direction of tuning is estimated per subject in crossvalidation, a procedure we describe in the supplementary methods. *Experiment 2 & 3*: power-law slopes of first-half test trials contained in the training distribution of that experiment's control condition, which are intermediate (35-65%) and all test trials (0-100%), respectively.

With this setup, we ran simulations where the representations were veridical (i.e. parameterized by the ground truth) as a baseline and compared these to simulations with: (1) early noise, i.e. noise at the pixel level, (2) late noise, i.e. noise in the location of the mean and (3) feedback noise, which does not affect inputs, but reflects lapses at the supervision signal level. This setup is detailed in Figure 2-4. For comparability, we used trial counts and a train/test structure exactly identical to that of human subjects, with the means of the Gaussians constrained by the learning curriculum. It was expected that both early noise and late noise, but not feedback noise, would selectively impair learning in the extrapolation condition. The rationale for this is visualized in Figure 2-5.

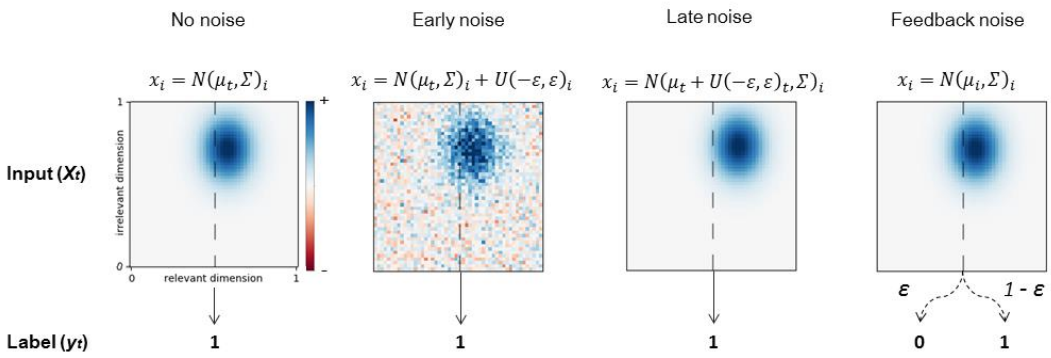

Figure 2-4: Condition overview of noise simulations. In this formalization, the learner's goal is to infer the label y from input X. $x \in X$ is the value of a single input pixel. $N(\mu,\Sigma)$ is the bivariate Gaussian distribution whose mean μ is a tuple composed of the trial's relevant and irrelevant parameter, and whose covariance Σ is fitted per subject. $U(min,max)$ is the uniform distribution and ε is the noise magnitude. Finally, subscript i marks pixel-unique terms, subscript t trial-unique terms, while terms without subscript were fixed for the entire experiment.

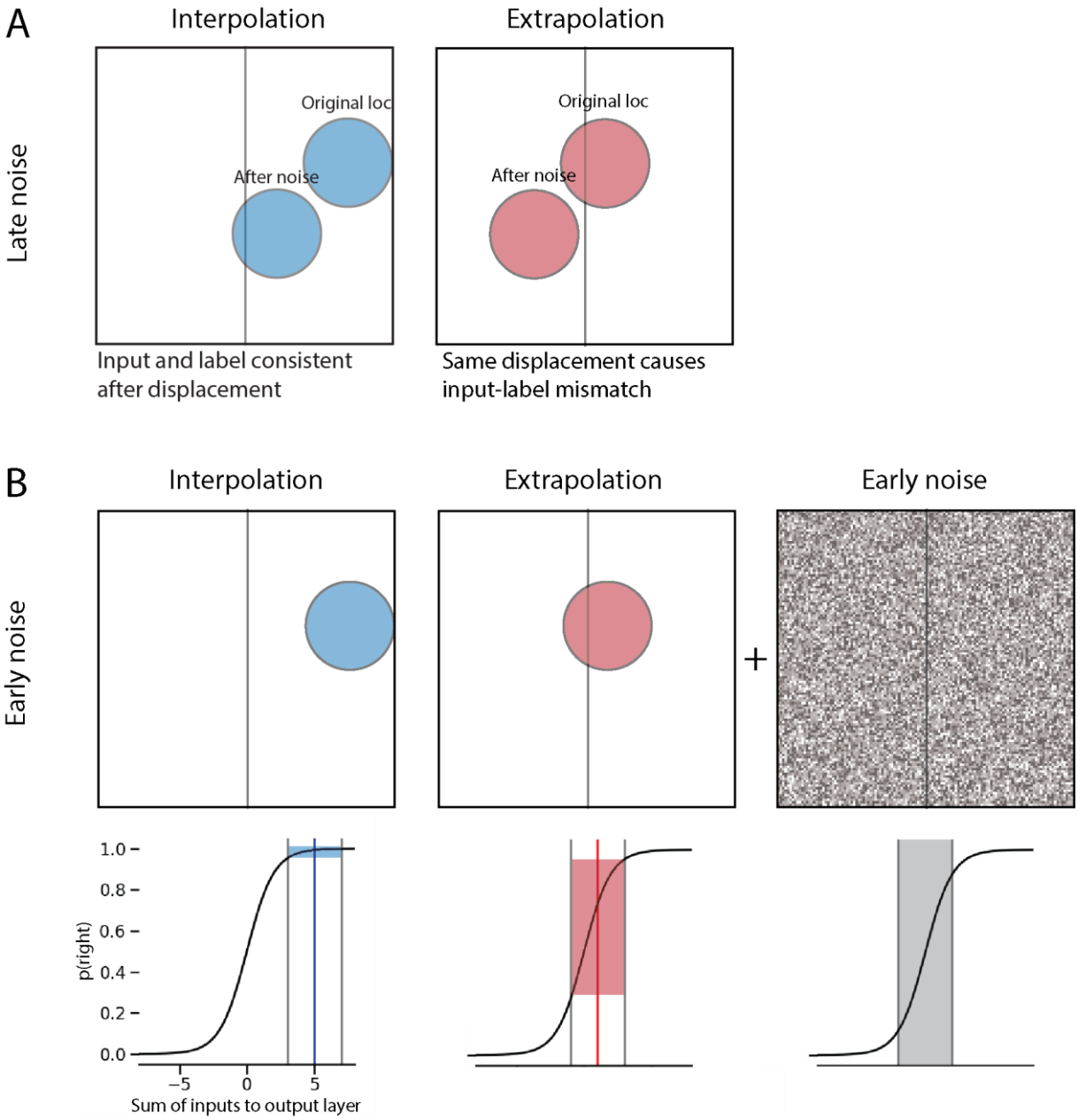

Figure 2-5: Intuition behind effects of noise. Conceptual illustration of how early and late noise may selectively impair learning in extrapolation. No condition-specific effects were expected for feedback noise. (**A**) When late noise is applied, a milder displacement is able to push extrapolation examples over the category boundary compared to interpolation examples, misaligning perceptual evidence with the received label. As in Figure 2-4, the x dimension represents the relevant dimension, the y dimension represents the irrelevant dimension, and the line through the center of the relevant dimension is the category boundary. Circles represent bivariate Gaussian inputs. (**B**) (**Top**): in the early noise condition, inputs are veridical, plus additive per-pixel noise. (**Bottom**) supposed effect on model output. The output layer contains a sigmoid nonlinearity (black). As all aspects of the task are symmetric, early noise is expected to contribute a symmetrically distributed additive noise term to the input of this sigmoid (gray). As extrapolation trial inputs are more ambiguous,

corresponding to a sum of inputs to the output layer close to 0, additive noise may produce large effects (red), whereas in the interpolation condition, the same amount of early noise may minimally affect model outputs, as the sigmoid is saturated (blue).

Demonstrated in Figure 2-6b is that the same noise type and magnitude can produce qualitatively different effects, depending on the curriculum in which they are applied. We observe that the effects of both early noise and late noise are most pronounced in extrapolation, and that this finding is robust to the choice of noise magnitude (Figure 2-6c). The time courses of their produced effects are different, however (Figure 2-6b). Unlike early and late noise, the effects of feedback noise are comparable between curricula.

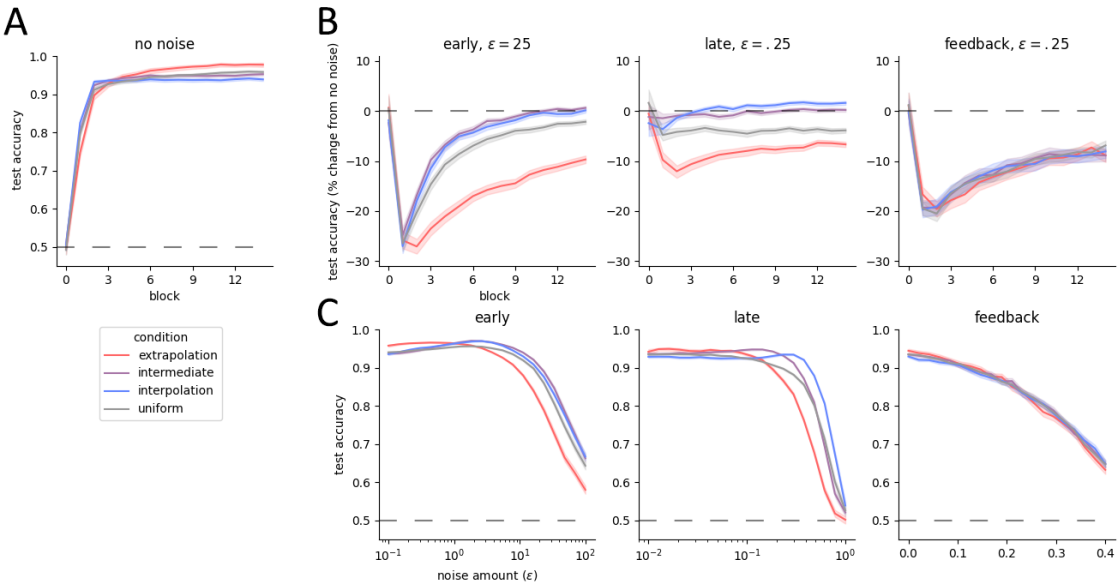

Figure 2-6: Simulated learning dynamics. (A) Per-condition baseline learning curves when no noise is used. (B) Per-block accuracy change relative to the no noise condition per noise manipulation. A fixed noise amount is used here for illustration purposes. (C) Mean overall test accuracy as a function of noise magnitude per noise manipulation. Error bars are 95% CIs over 500 cycles.

To establish how the simulations relate to human behavior, the abstract task model (neural network receiving bivariate Gaussian inputs) was fitted to human participants' responses in our main experiment, and compared using Bayesian model selection. Both fitting and testing methods are described in the supplementary methods. Human test trial responses were best explained by the late noise model, which had a protected exceedance probability (PEP) of 99.98%, crossing the 95% significance threshold. We note that this result was obtained by applying noise both at training and at

test, but that a consistent result is obtained when repeating the optimization pipeline and applying noise at training, but not evaluation, so that noise affects learning, but not inference (main effect PEP = 98.81%). An overview of per-model exceedance probabilities and model attributions in each condition is given in Figure a-3.

2.4.2 Neural correlates of learning

As electrophysiological correlates of learning, we leveraged two well-characterized EEG components. The first of these is the centro-parietal positivity (CPP), a marker of evidence integration or decision certainty (O'Connell, Dockree, & Kelly, 2012; Twomey, Murphy, Kelly, & O'Connell, 2015). The second is the feedback-related negativity (FRN), a post-feedback marker of prediction error (Holroyd & Coles, 2002; Huang & Yu, 2014; Pfabigan, Alexopoulos, Bauer, & Sailer, 2011). Mean CPP amplitude did not differ between conditions, $Z(46)=.8286$, $p = .407$. However, when predicting per-timepoint CPP amplitudes using psychometric regressors from a local time-window in a multiple linear regression model (details in the supplementary methods), a significant cluster was found for the relevant dimension regressor. Comparing mean beta coefficients between interpolation and extrapolation in the time-window corresponding to the relevant dimension main effect cluster, we find a significant difference in relevant dimension tuning, which is stronger in interpolation $Z(46) = -2.16$, $p = .031$. These effects are visualized in Figure 2-7 (Top). An exhaustive overview of per-condition regression weights, including all regressors and a wider time window is provided in Figure a-1, but contains no other significant clusters. For FRN, the traditional finding is a greater negative inflection after negative feedback than after positive feedback (Holroyd & Coles, 2002). Using a base-to-peak metric, we find a difference in FRN between correct and incorrect trials in interpolation, $Z(23) = 4.20$, $p < .0001$, but not in extrapolation, $Z(23)=.514$, $p=.607$. The difference of differences between groups is significant, $Z(46) = -3.143$, $p = .0017$. As FRN is feedback-driven, it was only computed for training trials. The FRN effects are visualized in Figure 2-7 (Bottom).

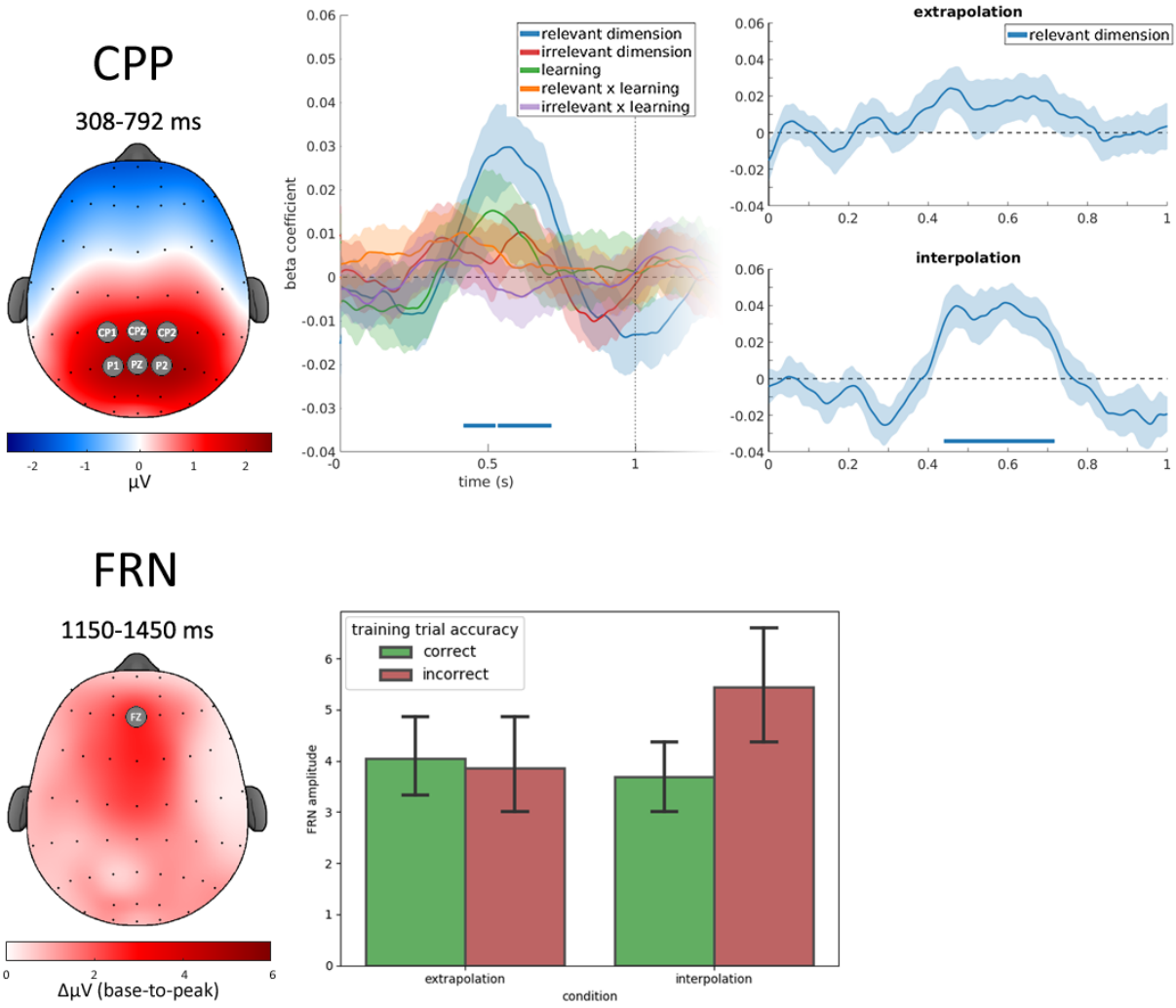

Figure 2-7: Electrophysiological results. (**Top Left**) Activation topography during CPP time window, time in milliseconds post stimulus onset. Highlighted electrodes make up the CPP EOI. (**Top Center**) Per-timepoint psychometric regression analysis. Plotted is time in milliseconds against regressor beta coefficients. The blue line indicates local relevant dimension tuning, while the red line indicates local irrelevant dimension tuning. The green line represents a power-law function fitted to per-subject accuracy. The orange and purple lines represent interactions between learning and the relevant or irrelevant dimension, respectively. Solid bars at the bottom of the graph indicate pointwise t-test significance at $p<.05$. Stimulus onset was at 0ms, and offset was at 1000ms. (**Top Right**) Relevant dimension effect in the same per-timepoint psychometric regression analysis, separated by condition. (**Bottom Left**) Topography of base-to-peak measures during FRN time window. (**Bottom Right**) FRN base-to-peak lengths for correct and incorrect trials, separated by condition.

2.4.3 Sequential effects

If learners benefit from an interpolation curriculum, this suggests a crucial role for exemplars with high relevant dimension extremity for achieving successful learning. The following analyses are aimed at better understanding the trial-to-trial dynamics of what happens when such exemplars are presented.

Using a generalized linear model (GLM) with probit linking function on behavioral test data, we examined how previous trial dimension extremity (i.e. unsigned distance from the category boundary) affected accuracy on the current trial, adding current trial dimension extremity as a competitive regressor. We then performed 2-sample t-tests on the obtained coefficients. Dimension extremity on the previous trial predicted accuracy more strongly on the current trial in interpolation than in extrapolation, $t(46) = 2.48$, $p = .0169$). In addition, coefficients were higher in interpolation for current trial relevant extremity, $t(46) = 4.88$, $p = .0155$, lower in interpolation for current trial irrelevant extremity $t(46) = -2.05$, $p = .0466$), while there was no effect of preceding trial irrelevant extremity $t(46) = .0155$, $p = .988$). These results are displayed in Figure 2-8a.

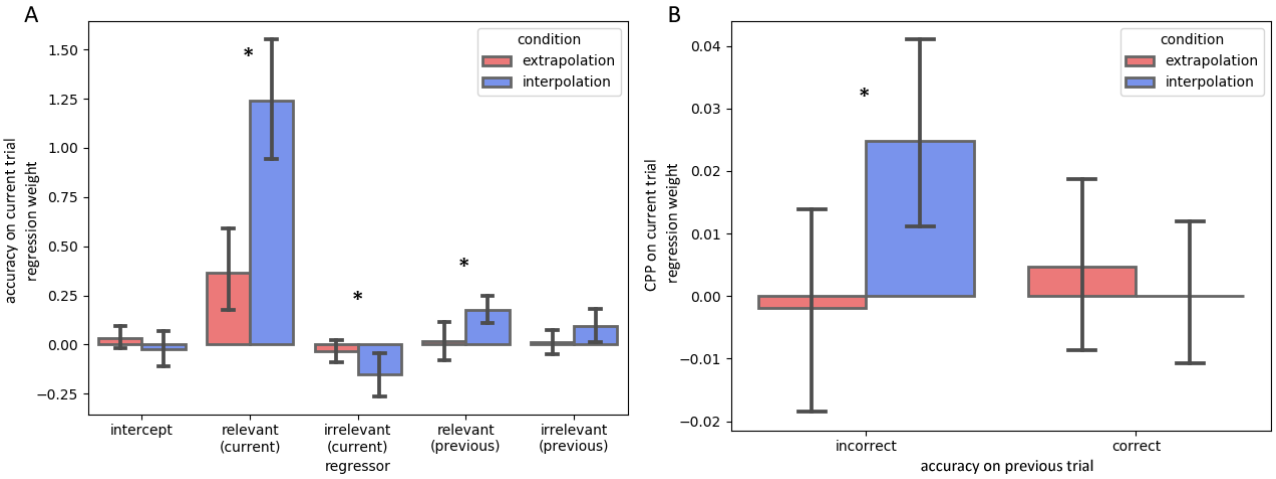

Figure 2-8: Sequential effects. (**A**) Behavioral sequential effects. Plotted are the beta weights of a GLM where accuracy was fit as a function of dimension extremity in relevant and irrelevant dimensions, both on the current and preceding trial. The different regressors are on the x-axis, while the y-axis represents the corresponding beta weights. (**B**) Neural sequential effect. GLM, regressing previous trial FRN amplitude against current trial CPP amplitude.

A final prediction was that learning should depend on examples that are objectively easy, but subjectively difficult, in line with Weinshall & Amir's theory (2020). Concretely, on an electro-

physiological level, we expected that only in the interpolation curriculum (objective simplicity), post-error FRN magnitude (subjective difficulty) would predict the amplitude of CPP in the subsequent trial. To test this, we used a GLM to fit subsequent trial CPP amplitudes as a function of training trial FRN amplitudes. Indeed, we find that the extent to which FRN amplitudes predict subsequent CPP magnitude is higher in the interpolation condition, $Z(46) = 2.46$, $p = 0.014$. This result is visualized in Figure 2-8b. As error trials become rarer over time particularly in the interpolation condition and this may affect the magnitude of prediction errors and hence FRN, we performed a control analysis, investigating whether FRN amplitudes for incorrect trials increase as a function of time in the interpolation condition. Using a t-test on per-subject correlation coefficients between test trial number and FRN amplitude, we find that this is not the case, $t(23) = -1.04$, $p = .310$.

2.5 Discussion

In this chapter, we sought to investigate how choice of learning curriculum can facilitate or impair learning of high-dimensional visual categories. We hypothesized that showing boundary-distal exemplars would bootstrap learning by guiding attention to diagnostic features. Conversely, difficult examples were expected to hinder learning by obscuring which differences among stimuli were category-relevant, and by increased sensitivity to noise. Our data confirms that learning was attenuated when early training exemplars lay proximal to the category boundary, while learning was facilitated when exemplars fell far from the category boundary. Furthermore, the performance benefit in the interpolation curricula transferred to proximal test exemplars in all three experiments, despite the fact that these exemplars were never encountered during training in the interpolation group, whereas the extrapolation condition was trained on them exclusively. Moreover, the interpolation curriculum improved performance on intermediate-difficulty trials compared to training exclusively on intermediate-difficulty trials, and outperformed a control condition which trained on all test trials equally when evaluated on all test trials. We conclude that training on distal exemplars confers a robust benefit on learning.

We then modeled curriculum learning in artificial learning systems, in order to disclose which computational mechanisms may underlie the observed behavioral effects. Using neural network

models, we incorporated limitations in human learning and decision-making as different types of noise. This revealed that learning is particularly sensitive to additive internal noise in the extrapolation setting, while interpolation confers a relative robustness to this noise. While this held true for both early and late noise, model comparison revealed that human behavior was best explained by late noise. An intuitive interpretation of this result is that in extrapolation, where proximity to the boundary is low, additive noise applied to percept centroids can more easily push an example across the category boundary, adversely affecting learning. Indeed, as visualized in Figure 2-6, the noise magnitude at which accuracy levels begin to decline are different between conditions, and ordered by distance to the category boundary, suggesting that distance-to-boundary acts as a buffer, protecting against deleterious effects of additive noise. This account may also explain why model attributions of the feedback noise model are higher in extrapolation than interpolation (Figure a-3): like late noise, this manipulation can cause perceptual evidence to become misaligned with feedback labels, late noise by shifting the percept over the category boundary on a proportion of trials, and feedback noise by inverting the label on a proportion of trials, yielding similar predictions. Early noise also produced curriculum effects qualitatively matching those observed in humans. Indeed, sufficient proximity to the category boundary can create ambiguous trials, with partial activation on either side of the category boundary. As such trials have an inherently lower signal-to-noise ratio, they are particularly sensitive to the further decrease in signal-to-noise ratio engendered by early noise.

Finally, we aimed to better understand the trial-to-trial mechanisms of learning through sequential analyses. In this regard, we report that increases in accuracy can be predicted as a function of previous trial distance-to-bound, suggesting a causal role of dimension extremity. In addition, a theoretical derivation suggests that the proportion of relevant dimension extremity over irrelevant dimension extremity should shape the quality of the angle of decision updates, while the magnitude of prediction errors should determine their magnitude. As predicted by this derivation, on the neural level, higher error-induced FRN was followed by higher amplitudes of CPP in interpolation only, establishing a correspondence between human learning and machine learning theory.

Whilst this work is primarily fundamental research, we expect our findings to have direct implications for educators. In particular, our findings suggest that ambiguous examples may be particularly harmful

(and vice-versa, curation helpful) in continued education, as the magnitude of late noise is known to increase with age (Yan et al., 2020). Our findings may also contribute to creating specialized learning curricula for clinical populations. For example, internal noise is greater in Alzheimer's Disease (Hogan et al., 2006) and schizophrenia (Yang et al., 2014). In ADHD, P3b is abnormally variable, which is pertinent as this implicates greater decision-stage noise (Saville et al., 2015). A particularly robust body of work demonstrates that internal noise is greater in those suffering from autism spectrum disorder (ASD), as inferred from behavior (Park, Schauder, Zhang, Bennetto, & Tadin, 2017), EEG (Milne, 2011) and fMRI (Dinstein et al., 2012). Even in subclinical populations, autistic traits are positively correlated with internal noise magnitude (Vilidaite, Yu, & Baker, 2017).

It is noteworthy that our behavioral findings contrast with those obtained in two prior studies on training schedule efficacy for medical images (Hornsby & Love, 2014; Roads, Xu, Robinson, & Tanaka, 2018). In these studies, no curriculum among easy-only, random training or an easy-to-hard schedule transferred to difficult examples. We attribute this apparent discrepancy to differences in stimulus selection and difficulty measures. Whereas our stimuli are algorithmically generated and difficulty can be traced back to a parametric setting in the generative process, the medical images and corresponding difficulty ratings were obtained empirically. Thus, there is no guarantee that these images vary along the same perceptual continua along the difficulty spectrum. Crucially, the most difficult empirical stimuli may include an unknown number of cases in which the perceptual evidence is misaligned with the formal diagnosis, which relies on additional information which is not available to the participant. In particular, in Hornsby & Love (2014), difficult examples were defined as those that were categorized below chance prior to training, and performance on these items remained below chance after training, regardless of training curriculum.

We also note that the terms early noise and late noise have been used to refer to somewhat disparate phenomena in prior work and that our treatment of these terms may differ slightly from their use by some authors. The crux of our perspective on these concepts is that early noise refers to a general degradation of the input signal, whereas late noise refers to a perturbation that is applied after some integration process, and so causes systemic deviations within learning episodes.

In future research, it would be valuable to explicitly investigate the role of dimensionality in determining curriculum success, for example with a modelling approach similar to ours, which expands to a greater number of dimensions. In addition, attention is likely to play a crucial role in any noise-based account of learning, as attention can suppress internal noise (Lu & Dosher, 1998), and the allocation of attention over stimulus dimensions is nonstatic over the course of learning (Rehder & Hoffman, 2005), with learners sequentially attending different dimensions based on their current hypotheses (Akaishi, Kolling, Brown, & Rushworth, 2016). This suggests a rich interplay between how learners divide their attention over dimensions, and how they are affected by curriculum and dimension extremity.

Chapter 3: probabilistic evidence integration: curricula and capacity limits

3.1 Chapter Abstract

Many of life's decisions, such as deciding the best destination for a vacation, or who to bet on for a boxing match, require integrating the information from multiple sources of information. When outcomes have multiple predictors, learners must solve a complex credit assignment problem in order to update their beliefs. Across 7 experiments, we tested how probabilistic evidence integration problems are learned and solved.

Our first conclusion is that humans are remarkably adept at overcoming the credit assignment problem, learning per-cue outcome probabilities equally well when these must be extracted from multi-cue problems as when trained on single cues. Despite this, they perform better at single-cue problems than multi-cue problems after either training method, suggesting that their ability to integrate over multiple cues to make correct decisions is limited.

When queried on multi-cue problems following either training method, no accuracy differences between groups were observed. Although full feedback was provided and the single-cue groups had no prior exposure to the correct integration principle, accuracy did not improve over the course of this multi-cue training.

To explain these findings, we adapted a current model from the literature which assumed seamless cue integration to instead use fast and frugal heuristic decision-making strategies. The first heuristic considered is the tallying heuristic, in which the predictive direction of cues is used, while cue validities are ignored. The second heuristic we modelled is the take-the-best heuristic, in which decisions are based on a subset of cues with high validity, while low-validity cues are discarded. Fitting these models to human data, the take-the-best model provided the best overall explanation of the data in each

experiment, improving upon the base model. One exception to this was formed by a curriculum which contrasted cues of equal validity, but opposite sign. Here, data were best explained by the tallying model, suggesting that the properties highlighted by training curriculum affect subsequent decision strategy.

Taken together, these results suggest that probabilistic evidence integration in human subjects is subject to a capacity limit. More generally, the work showcases that heuristics have an important role not only in the decision-making literature where they are normally encountered, but also affect learning, providing an exciting avenue for future research.

3.2 Introduction

When stepping out for a brief walk, one can either bring their umbrella, or choose not to. How are such decisions made? Even for seemingly simple decisions, a wealth of information sources are available. These sources may provide only partial information, such as the amount of light shining through the drapes when you woke up. Depending on your level of trust in meteorology, a forecast of heavy downpour may be deemed more reliable, although even this may be overturned upon discovering clear skies and shining sun. This example illustrates several challenges in probabilistic learning settings. The learner has to identify which sources of information are relevant for predicting outcomes. Then, they must integrate the information from these different, potentially conflicting sources. Once an outcome is obtained, the learner faces a credit assignment problem where they must decide how an observation changes their beliefs about the credibility of the observed predictors for future choices.

In a well-known formalization of this type of problem known as the weather prediction task (WPT; Knowlton, Squire, & Gluck, 1994; Yang & Shadlen, 2007), learners are faced with combinations of cues, a limited subset of which is drawn on any given trial. Each cue bears a probabilistic relationship to binary trial outcomes, corresponding to a likelihood ratio of reward given either response. The correct integration rule is to sum up the log-likelihood ratios of all present cues, often called WOE (weight of evidence) in this context. When cues are presented sequentially, neural signals follow this normative process of per-cue addition or subtraction, as observed in monkey LIP neuron firing rates (T. Yang & Shadlen, 2007) and human beta-band activity (Gould, Nobre, Wyart, & Rushworth, 2012). The

majority of work on this topic has considered the processes engaged when subjects pre-trained to asymptotic performance perform this task (Gould et al., 2012; Yang & Shadlen, 2007) and how probabilistic learning relates to implicit and explicit memory (Knowlton et al., 1994; Price, 2009). A comparatively smaller body of work has examined how subjects learn rather than execute the task (Akaishi et al., 2016; Gluck, Shohamy, & Myers, 2002). To our knowledge, none have investigated the determinants of learning success.

The original hypothesis motivating the experiments in this chapter was that subjects may benefit from training on individual cues in learning to solve the multi-cue problem, carving a complex combinatorial problem at its joints. Indeed, the close match between neural mechanism and normative solution suggests that the integration principle is learned naturally. Moreover, work investigating learning strategies shows that when subjects are trained on a variable number of cues, they initially choose to focus on trials where only a single cue is presented, transitioning to more complex strategies only later on (Gluck et al., 2002). Taken together, this seems to suggest that the bottleneck in probabilistic information integration learning is in learning the correct cue weights through appropriate credit assignment, rather than learning the integration rule. In accordance with this view, in the most complete current process model of the WPT, learning consists only of updating cue weights, whereas the integration rule is hardwired and lossless (Kurzawa, Summerfield, & Bogacz, 2017).

Across 7 experiments, we found that subjects trained on a variety of curricula indeed showed learning benefits during training, suggesting knowledge of the cue weights. However, to our surprise, this knowledge robustly failed to transfer to the multi-cue test setting. Shifting focus to the mechanisms that may explain this pattern of behavior, we adapted the current model (which assumes seamless cue integration) to use suboptimal heuristic integration strategies observed in decision-making settings.

Such “fast and frugal” heuristics were popularized by Gigerenzer and colleagues (Gigerenzer & Selten, 2002; Gigerenzer & Todd, 1999), who framed the mind as an “adaptive toolbox”, consisting of simple decision rules, each fit for a different set of problems and settings. Humans’ adaptive success despite limited resources is then interpreted as competence in selecting the right tool in the right context. This allows humans to flourish despite being only “boundedly rational” (Simon, 1957) – rational only when

accounting for processing constraints and the costs of gathering information. Later work has shown that how subjects choose from this toolbox, i.e. which heuristic decision rules they employ, depends on a multitude of factors, including availability of information on cue rankings and the predictive direction of cues (Binz, Gershman, Schulz, & Endres, 2020) use of visual versus verbal cues (Bröder & Schiffer, 2003), cue salience and cue validity (Bryant, 2014), linearly versus non-linearly separable problems (Hoffrage, Garcia-Retamero, & Czienskowski, 2005) and investment costs for obtaining information (Bröder, 2000).

In particular, we focus on the tallying heuristic (Czerlinski, Gigerenzer, & Goldstein, 1999; Dawes, 1979), in which cue representations are binarized to whether they are positive or negative predictors, after which decisions are based on the sign of their sum. We also examined the take-the-best heuristic (TTB; Gigerenzer & Goldstein, 1996), in which cues are first ordered by their validity, and only a subset containing the most valid cue(s) is used. After introducing the general methodology, the empirical results will be detailed per set of experiments, followed by the modelling analyses, which will involve data from all experiments.

3.3 General methods

3.3.1 Participants

A total of 525 participants took part in the 7 experiments described here. All participants gave informed consent and read through written instructions before starting the task. All experiments were approved by the Medical Sciences Research Ethics Committee of the University of Oxford (approval reference: R50750/RE001). Participants were compensated at an approximate hourly rate of £10. Participants in experiment 2 completed the task in-lab in the psychology department of the University of Oxford. The data for the other experiments were collected online via https://www.prolific.co. An age restriction of 18-40 years was used. For data quality, subjects were also required to have an approval rate of at least 85% over at least 5 previous studies on the platform.

Subjects were excluded from the analyses if they gave the same response for more than 25 trials in a row, or when they responded faster than reasonably required for processing the stimuli, defined as

responding in <300 ms on at least 150 of trials in experiments with free responses. This cut-off was based on experiment 2, where the mean plus 3 standard deviations of this metric was at 149 trials. For experiments with a response lockout, pre-stimulus onset responses were measured instead, and subjects were similarly excluded when these occurred in at least 150 trials and numbered at least 3000 responses, indicative of a repetitive response strategy.

3.3.2 Design and procedure

Stimuli were two sets of 8 differently colored shapes, either geometric shapes or stylized animals. One set was used per subject, with the choice of stimulus set counterbalanced between subjects. Stimuli consisted of a maximum of 4 cues, presented per trial in a 2x2 array occurring in the centre of the screen, where cue location was irrelevant for trial outcomes. Each experiment consisted of a training phase and a test phase, each of 400 trials (8 blocks of 50 trials). In the test phase, each stimulus comprised 4 cues. Our experimental manipulation is the curriculum assigned during training. In the basic "curriculum" condition (Exp1-4 and 7), training stimuli comprised a single cue. In the other experiments, variations of this curriculum were used, but always contained singletons (trials with a single-cue stimulus) to emphasize individual cues. The "parallel" condition is similar in all experiments, and forms a control condition in which training is performed on the test distribution, generally involving 4 cues per trial. A full overview of the particulars of each experiment is given in section 3.3.3, Table 1.

Each cue carried probabilistic information about the trial outcome. Outcomes were good or bad weather, randomly assigned to the left and right buttons per subject, and indicated by a rain cloud and sun, respectively. In all experiments, the same set of probabilities, [.1, .2, .3, .4, .6, .7, .8, .9], was assigned randomly to the 8 cues. The WOE of a cue *i* is defined as the log-odds corresponding to its probability p_i, framed here in the frame of reference of the right button being correct:

$$WOE_i = log(\frac{p_i}{1-p_i}) \quad \text{(Eq. 3.1)}$$

Then, the integrated probability $p(reward|right)$ given a set of present cues $\{p_1, p_2, \ldots, p_m\}$ is:

$$p(reward|right) = \frac{e^{WOE_{sum}}}{1 + e^{WOE_{sum}}},\ with\ WOE_{sum} = \sum_{i=1}^{m} WOE_i \qquad \text{(Eq. 3.2)}$$

The cue-to-probability mappings remained constant throughout the training and test phase for each participant. Cues were randomly sampled, without replacement in Exp1-5, and with replacement in Exp6-7. As a constraint, stimulus combinations that resulted in *p*(reward | right) = *p*(reward | left) = .5 were not used. After making their prediction, participants received feedback. Feedback was provided throughout training and test in all experiments. As feedback was probabilistic, the received feedback could deviate from the normatively correct answer. The type of feedback was indicated by a change in color of the fixation dot (green when positive, red when negative) and an accompanying sound effect. Trials began with fixation, which was displayed for 500ms, followed by stimulus onset. In experiments with a response lockout, the response lockout lasted for 2s. After this, the response window opened, indicated by a change in color of the fixation dot from black to gray. Cues remained on-screen during the 20-second response window (a countdown progress bar appeared 5 seconds before the end of the window). After response or timeout, the stimulus disappeared, and the fixation dot changed in color to indicate feedback. Feedback was displayed for 700 ms when positive and 1000 ms when negative. This trial structure was used for both training and test trials. The trial structure is visualized in Figure 3-1.

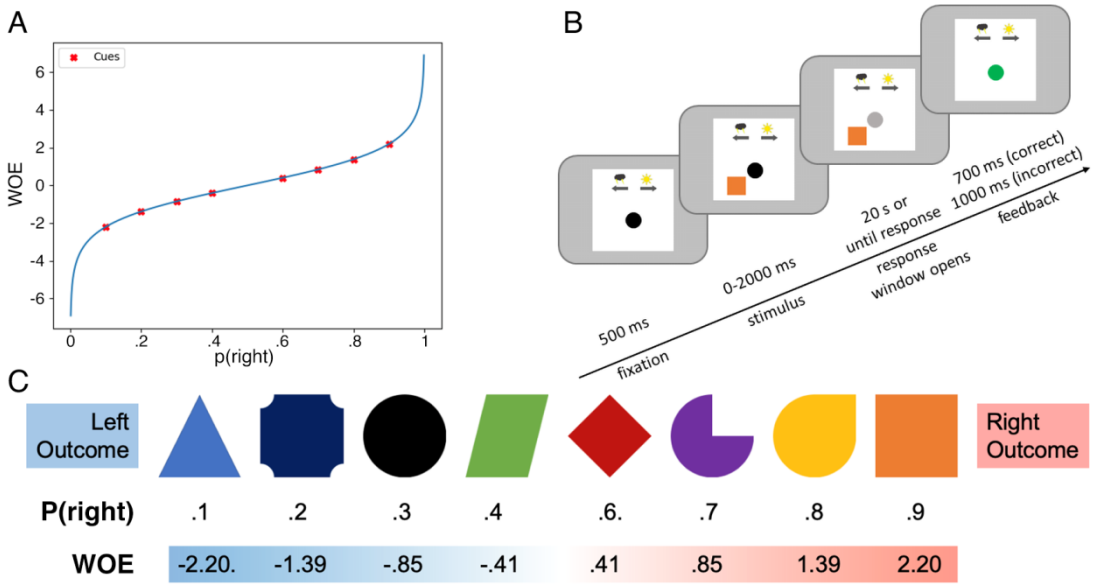

Figure 3-1: Task setup for all experiments. (**A**) Probability that the rightward response was rewarded plotted against WOE. The locations of the 8 cues are indicated in this space as red points. (**B**) Trial structure for both training and test trials. Fixation dot size is increased for visualization purposes, otherwise to scale. Fixation was displayed for 500ms, after which the cues appeared. In experiments with a response lockout, this was followed by a 2s pause. After this, the response window opened, indicated by a change in color of the fixation dot from black to gray. After response or timeout, the stimulus disappeared, and the fixation dot changed in color to indicate feedback. For incorrect responses, the feedback was displayed for an additional 300ms. (**C**) Example stimuli with corresponding probability of rightward response being rewarded and WOE. Assignment of shapes to probabilities was randomized. The same probabilities were used in all experiments.

The main task started after participants read through written instructions with accompanying example images (but no information on per-cue outcome probabilities) and successfully completed a brief warm-up. The warm-up was a simpler version of the main task, using 2 cues with a p_i of 10% and a 90% probability cue, respectively. Single- cue stimuli were presented for 20 trials, and the warm-up was repeated until subjects gave the normatively correct response on at least 15/20 trials. The stimuli used for the warm-up were always a trumpet and a guitar, to prevent interference with the main task.

3.3.3 Logistic regression model

For inferring the empirical weight given to cues in participants' choices, a logistic regression model was used. This model was fitted separately at training and test for each participant. The model estimates the probability of the rightwards response as a function of the number of times a cue was present in a trial N_i, per-cue weights w_i and an intercept w_0:

$$\hat{p}_{right} = \sigma(w_0 + \sum_{i=1}^{8} w_i N_i) \quad \text{(Eq. 3.3)}$$

Where $\sigma(x)$ is the canonical sigmoid function:

$$\sigma(x) = \frac{1}{1 + \mathrm{e}^{-x}} \quad \text{(Eq. 3.4)}$$

Parameters were estimated using the Broyden–Fletcher–Goldfarb–Shannon algorithm. L2-regularization was used as the optimization target is otherwise underdetermined – if the same number of cues is presented on every trial, an infinite number of equivalent solutions can be obtained by trading off intercept and per-cue weights. For statistical tests comparing weight magnitude between groups, the mean weight per participant is used, ignoring the intercept, and reversing the sign of weights corresponding to cues with $p_i < .5$:

$$w_{tot} = \frac{1}{8}\sum_{i=1}^{8}(w_i \cdot sign(WOE_i)) \quad \text{(Eq. 3.5)}$$

We refer to this quantity as "total weight". The sign adjustment ensures that weights contribute a positive value when they are in the normatively correct direction, and a negative value when they are in the incorrect direction.

3.3.4 Experiments

While full details on the individual experiments will be given in their respective sections, for clarity we provide an overview here with the most salient details of each experiment.

Experiment name	Test accuracy Median (IQR)		Mann-Whitney-U on test accuracy difference	N after (before) rejection	Curriculum description
	After Curriculum training	After Parallel training			
Exp1: Pilot	.61 (.53-.65)	.56 (.53-.61)	*U*=154, *p*=.049	44 (54)	Exactly follows general methods
Exp2: In-lab replication	.66 (.58-.70)	.65 (.60-.69)	*U*=414, *p*=.381	60 (64)	In-lab replication of Exp1
Exp3: 2-stim integration	.69 (.61-.73)	.69 (.64-.72)	*U*=499, *p*=.446	66 (80)	2 stims/trial in parallel training trials and all test trials
Exp4: Integration training	.61 (.54-.73)	.55 (.50-.64)	*U*=548, *p*=.297	75 (92)	Alternating train/test blocks, otherwise identical to Exp1.
Exp5: Blocked stim identity	.54 (.51-.58)	.61 (.55-.68)	*U*=148, *p*=.007	54 (102)	Curriculum: blocked training on cue pairs (Parallel unchanged)
Exp6: Comprehensive curriculum	.73 (.66-.76)	.71 (.61-.74)	*U*=655, *p*=.114	79 (84)	General design changes in exp 6-7: - 2s response lockout - Stimuli sampled with replacement Exp6 Curriculum: - Stimuli introduced in pairs in descending order of validity. - Blocks alternate singleton training on new cues with integration training on growing pool of introduced stimuli
Exp6b: Sequentially built up arrays	.72 (.69-.77)	-	-	36 (40)	As Exp6, but cues appeared in the array one-by-one, with a 500ms delay between each cue.
Exp7: Exp 1 + non-feedback singletons at test	.64 (.56-.69)	.69 (.63-.71)	*U*=173, *p*=.081	45 (50)	As Exp1, with non-feedback singleton trials added during the test phase. General design changes from Exp6 were retained.

Table 3-1: Overview of experiments. The first column indicates the experiment numbers. Columns 2-3 show per-group performance on test trials. Reported are medians, followed by first and third quartiles in brackets. Column 4 provides test statistics and *p*-values of Mann-Whitney-*U* tests comparing parallel test trials accuracy between groups. Column 5 displays the post-rejection subject count, with the pre-rejection count in brackets. Column 6 provides a brief description of how each experiment deviated from the basic design described in the General methods.

3.4 Experiment 1-2: training on single-cue vs. multi-cue trials

Our first experiment followed the general design without any deviations. Experiment 2 is a replication of experiment 1, performed on-site at the Department of Experimental Psychology of the University of Oxford. Their aim was to test the hypothesis that training on singleton trials improves accuracy on 4-cue parallel test trials relative to training directly on the test distribution. In order to gauge whether there were differences between performance on observed cue combinations and generalization to novel cue combinations, in the parallel condition of experiment 2, we held out half of the valid cue combinations (32 combinations) during the training phase. The set of observed combinations was constrained to feature each cue in at least 2 separate combinations.

3.4.1 Results

In experiment 1, this prediction was borne out and test accuracy was higher in the curriculum condition (N = 22, median = .61) than in the parallel condition (N = 20, median = .56), although barely meeting the significance threshold, $U = 154$, $p = .049$. This did not replicate in the lab in experiment 2, where no difference in test accuracy was found between curriculum (N = 30, median = .66) and parallel (N = 29, median = .65), $U = 414$, $p = .381$. Learning curves for training and test are displayed in Figure 3-2a. In experiment 2, no difference was found in accuracy on held-out versus previously observed cue combinations in the first test block $T=201$, $p=0.721$. Only the first test block was considered, as full feedback was provided during test, providing learning opportunities for the previously unobserved cue combinations.

Using a logistic regression model to estimate the weight each cue had on subject choice in experiment 2, total weights were higher during training (median = 1.19) than in test (median = 0.81) in the curriculum group, Wilcoxon signed-rank test $T = 38$, $p = .0001$, while the converse was true in the parallel group (training median = 0.58, test median = 0.73),$T = 69$, $p = .0013$. At training, total weights were higher in the curriculum condition than in the parallel condition, $U = 77$, $p < 0.0001$, while at test, total weights were not significantly different between conditions, $U = 428$, $p = .461$. Results of the logistic regression analysis are visualized in Figure 3-2b.

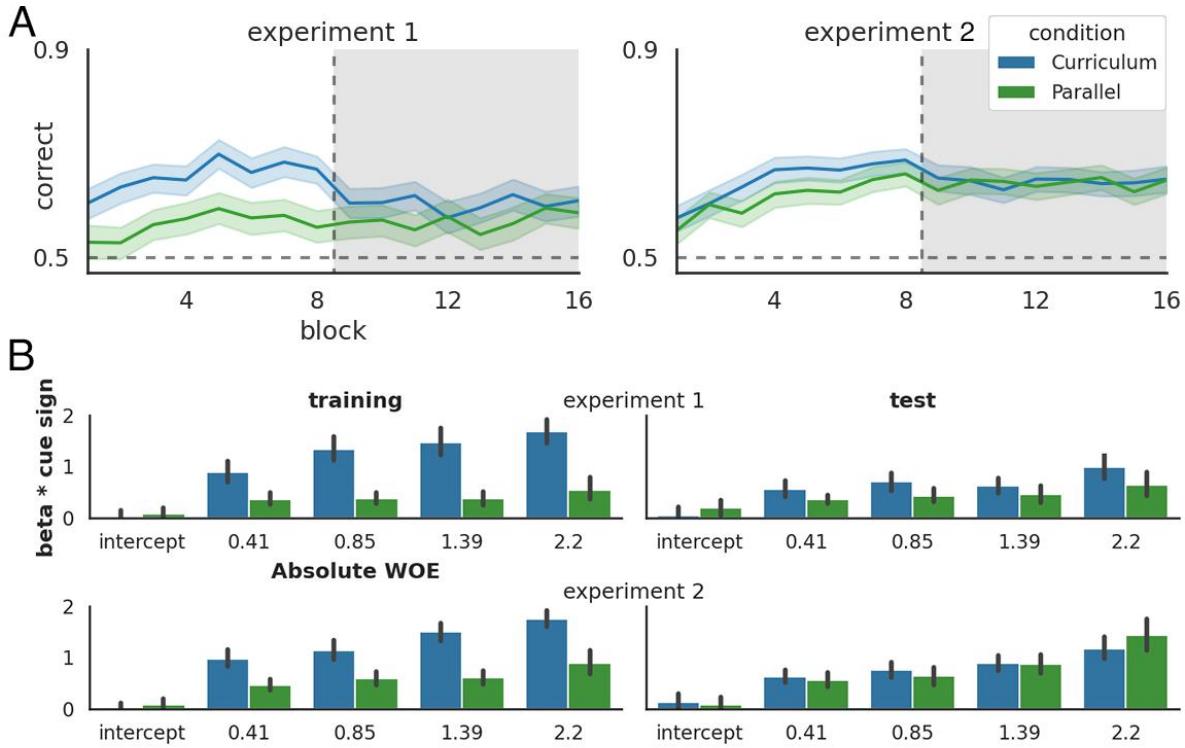

Figure 3-2: Results in experiment 1 and 2. The curriculum condition is displayed in blue, and the parallel condition in green. (**A**) Learning curves, displaying fraction of correct trials against experimental block. Shaded region indicated test phase. Horizontal dashed line indicates chance level. (**B**) Outcome of logistic regression, which models the probability of a rightward response as a function of the cues present in a trial. Displayed are the intercept and per-cue weights, grouped by absolute WOE. For cues with a negative WOE, the sign of the obtained weights was reversed to make them comparable to their positive-WOE counterparts. Learning curves and logistic regression weights are provided for all other experiments in section 3.8.

The replication failure of the test accuracy difference between conditions makes it likely that the weak effect obtained in experiment 1 is a type I error. While these findings suggest that subjects in the curriculum learn appropriate cue weights and utilize these during training, they raise the question why this does not yield any differential benefits at test. There are several possible reasons why this setup may not have been conducive to integration success in the curriculum group. First, presenting 4 cues at the same time makes the number of possible combinations large, likely discouraging rote learning. Potentially, generalization benefits of the curriculum would be more pronounced if rote learning were a feasible strategy in the parallel condition. Alternatively, withholding any form of integration until the onset of test may have had deleterious effects on learning, as subjects were not required to learn cue

validities (i.e. their absolute WOE) during training, only their signs, while cue validities were essential during test. Finally, as feedback was probabilistic, learning accurate per-cue estimates involves integration over many trials. In this regard, it has previously been demonstrated that interleaving the different units over which integration must take place (as we did with cues) fosters interference, whereas blocked schedules promote disentanglement (Flesch et al., 2018). These three concerns will be addressed in the next set of experiments.

3.5 Experiment 3-5: determinants of integration success

3.5.1 Motivation & Methods

The first set of experiments suggested that while training on singleton stimuli improved effective per-cue weights on these trials, this did not transfer to 4-cue integration trials, compared to training on the latter directly. To account for the possibility that presenting 4-cue combinations discouraged rote learning, in experiment 3, we used only 2 cues on integration trials. The motivation behind this manipulation is that the limited number of non-chance 2-cue combinations can feasibly be learned through rote memorization, but this should hinder generalization if not all combinations are observed. To test this, half of the possible combinations of cues (14 combinations) were held out during training in the parallel condition, allowing us to gauge whether performance on these differed from performance on observed combinations.

In experiment 4, to test whether exposure to the subsequent integration task would benefit curriculum participants, for example by alerting them to the fact that there was merit in learning the validity of each cue, we alternated training and test blocks (50 trials each), instead of presenting all 8 training blocks before proceeding to test.

In experiment 5, we blocked cues during curriculum training by restricting the number of cues which could appear in any given block. The most stringent blocking design would be to allow only a single type of cue appear in each block, one cue type for each of the 8 blocks. However, this was deemed impractical, as participants would not have to process the stimuli to do well during training, and the desired behavior would be monotonous: to press the same button for all trials in a block. Instead, we

limited training trials to 2 different cues for every set of 2 training blocks. On every trial, these were either presented as singletons, or as 2-cue combinations, featuring each cue once. The order in which cues were introduced was random, with the constraint that the second cue could not have the same validity as the first cue, which would otherwise turn their combination into chance trials. The parallel condition remained unchanged.

3.5.2 Results

In experiment 3, in the parallel group, accuracy in the first test block was higher on examples shown during training than on examples withheld until test, Wilcoxon signed-rank test $T = 116$, $p = .017$. This suggests that subjects in the parallel condition did rely to some extent on rote memorization to solve the task. Only accuracies in the first test block were used, as full feedback was provided during test, providing opportunities to learn the previously held-out examples. Overall test accuracy was not significantly different in the curriculum group (N = 34, median = 0.69) compared to the parallel group (N = 30, median = 0.69), $U = 499$, $p = 0.446$.

In experiment 4, test accuracy was not significantly different in the curriculum group (N = 40, median = 0.61) compared to the parallel group (N = 38, median = 0.55), $U = 691$, $p = 0.247$, suggesting that interleaving training and test blocks did not produce overall curriculum effects.

In experiment 5, test accuracy was unexpectedly lower in the curriculum group (N = 16, median = 0.55) compared to the parallel group (N = 33, median = 0.61), $U = 148$, $p = 0.007$, suggesting that introducing the cues in blocked pairs had harmed rather than helped learning. The exclusion rate was very high in this experiment (47% of subjects were excluded, while previous experiments had exclusion rates of 18-21%), primarily due to sub-criterion reaction times. As these disproportionally affected the curriculum group, it is likely that a proportion of subjects discovered that due to the blocking structure, they did not have to process the stimuli to predict trial outcomes in some blocks. With this in mind, we also report the result of the same test without applying exclusion criteria. In that case, no difference is obtained between the curriculum group (N=54, median=0.54) and the parallel group (N=48, median=0.56), U=1067, p=0.063.

Taken together, we conclude that none of the manipulations in this section succeeded in promoting integration success in the curriculum group past the level of their respective parallel control groups. However, of note is that generalization to unseen cue combinations was impaired relative to performance for seen combinations in the 2-cue integration setting of Exp3. As this effect was not obtained in the 4-cue integration setting of Exp2, the scalability of this result is likely bounded by the feasibility of rote learning the available cue combinations.

In the next experiment, which is our final attempt to promote probabilistic information integration learning, we slightly overhauled the general design to allow us to use a more comprehensive curriculum, and to prevent the high attrition rates obtained in experiment 5.

3.6 Experiment 6: comprehensive curriculum

3.6.1 Motivation & Methods

In the introduction of this thesis, it was described that time functions as an important scaffold for learning. Recognizing this, in experiment 6, we introduced cues in a meaningful order – their validity. Furthermore, between blocks we alternated between introducing a new set of cues using singleton trials, and integration training, using only the pool of introduced cues. The motivation for this is to provide spaced repetition and an opportunity to learn the integration rule in the integration blocks, whilst making it straightforward to learn about the individual cues in the singleton blocks. A full overview of experiment 6's curriculum is given in Table 3-2.

In addition, two general design changes were made. As mentioned in the previous section, integration trials involving a cue and their opposite-valence counterpart are problematic in the setting without replacement. For this reason, we switched to sampling with replacement. Furthermore, as the previous experiment had a high exclusion rate based on infeasible reaction times (less than 300ms), we added a response lockout of 2 seconds from stimulus onset.

Block		1	2	3	4	5	6	7	8	Test (9-16)
Curric-ulum	Cues/Trial	1	4	1	4	1	4	1	4	4 cues/trial sampled from all cues
	Cue pool	.9, .1	.9, .1	.8, .2	.9, .1, .8, .2	.7, .3	.9, .1, .8, .2, .7, .3	.6, .4	All cues	
Parallel	Cues/Trial	4								
	Cue pool	All cues								

Table 3-2. Curriculum structure in experiment 6. The curriculum alternates between introducing new cues and integration training on the pool of introduced cues. Cues were introduced in pairs, in descending order of validity. In the parallel condition, only 4-cue trials were used during the training phase, as in all other experiments.

3.6.2 Results

Test trial accuracy in the curriculum group (N = 41, median = 0.73) was not significantly different from test accuracy in the parallel group (N = 38, median = 0.71), U = 655, p = 0.114, suggesting that the comprehensive curriculum did not promote learning relative to control.

3.6.3 Sequential presentations control (Experiment 6b)

As previous studies that link the processing mechanism to the normative procedure of adding up per-cue WOEs employ sequential presentations (Gould et al., 2012; Yang & Shadlen, 2007), we collected a control condition to investigate whether sequentiality is fundamental to integration success. All details were as in the curriculum group, with the exception that instead of appearing all at once, stimuli came on one at a time every 500 ms, gradually building up the array. However, we found no effect of sequential presentations (N = 36, median = .72) vs. immediate presentations (N = 41, median = .73), U = 723, p = .439

3.7 Experiment 7: Individual cue estimates

3.7.1 Motivation & Methods

Section 3.4 showed that curriculum subjects were more accurate during training compared to their parallel counterparts, but that the behavior of both groups was indistinguishable on the subsequent parallel test. However, it is not unequivocal that this implies that both groups employ qualitatively similar solutions for 4-cue trials. Alternatively, it is possible that both types of training schedules emphasize different parts of the task, such as individual cue contributions in the curriculum condition and integration in the parallel condition, so that the relative advantages and disadvantages of these training schedules cancel out in their resulting test scores. Given the consistency of our results so far, we deemed it more likely that subjects in the parallel conditions had succeeded in learning about the individual cues, and that the apparent differences in behavior between conditions during the training stage were instead a result of limitations in integration capacity. If so, parallel participants' performance on singletons should not deviate from that of participants trained only on singleton trials. We tested this prediction in the current experiment. For this purpose, we reverted to the design of experiments 1 and 2, where the only difference between curricula was the number of cues displayed during training trials. To evaluate our prediction, we added interspersed singleton trials in the test phase, for which no feedback was provided. Neutral feedback was indicated by the dot turning a lighter shade of gray and was presented for 850ms, mid-way between the 700ms used for correct responses and the 1000ms used for incorrect responses. Each cue was presented twice per block in this fashion, for a total of 16 additional singleton trials per test block. In addition, we added a post-experiment questionnaire, in which participants were asked to indicate the probability of good weather for each cue, and their confidence in this estimate, both as a percentage between 0 and 100, indicated by a slider. Questions were framed in the frame of reference of the rightward response option, and the good and bad weather options were displayed on the same sides as in the main experiment. An example screenshot from this questionnaire is displayed below in Figure 3-3.

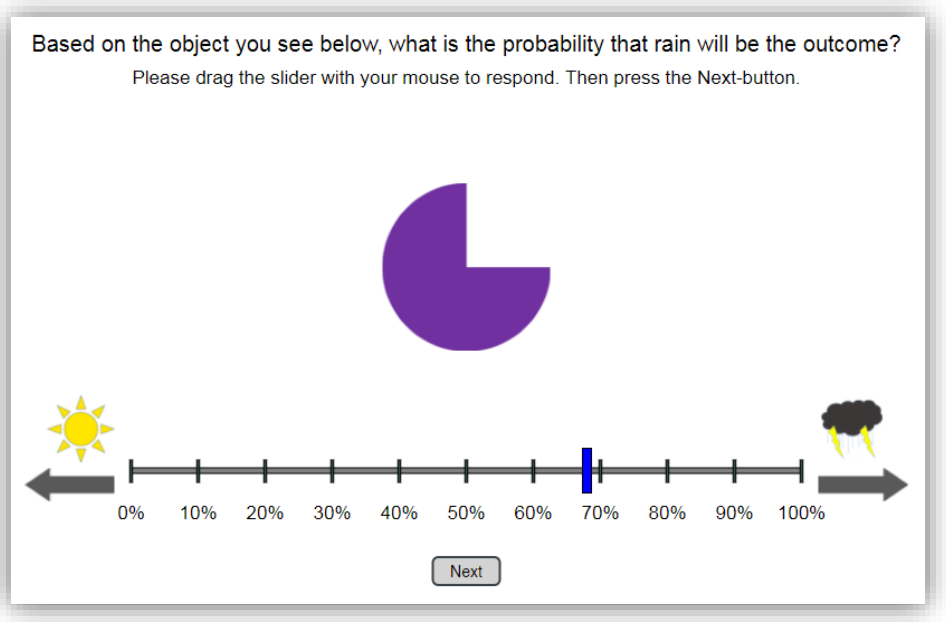

Figure 3-3: Post-task questionnaire. After the main task in Experiment 7, subjects were asked to indicate the outcome probability for each cue, and their confidence in each rating.

Our predictions were that performance on singleton test trials would exceed performance on 4-cue test trials in the parallel group, and that singleton test performance would not differ between groups. As the changes of using a response lockout period and sampling with replacement introduced in experiment 6 were considered to be general methodological improvements, these changes were retained in the current experiment.

3.7.2 Results

Performance on 4-cue test trials was not significantly different in the curriculum group (N = 21, median = .64) compared to the parallel group (N = 22, median = .69), $U = 173$, $p = .081$. For test-time singleton trials, there was no difference in accuracy between the curriculum group (median = 0.75, IQR = .60-.86) and the parallel group (median = .83, IQR = .66-.89), $U = 212$, $p = .18$. In both groups, singleton trial test performance was higher than parallel trial test performance (Curriculum group $T = 233$, $p < 0.001$. Parallel group $T = 233$, $p = 0.009$).

In the post-task questionnaire, mean absolute errors in outcome probability estimation over all cues were not different between curriculum (median = 0.18, SD = 0.07) and parallel (M = 0.19, SD=0.10), U=228, p=0.476. Subjective confidence ratings were also not different between curriculum (median = 0.69, SD = 0.12) and parallel (median = 0.70, SD = 0.18) conditions, U = 209, p = 0.301.

In conclusion, our main prediction for this experiment was confirmed: learners succeeded at extracting per-cue weights when faced with a parallel training schedule that involved a credit assignment problem, being unimpaired relative to subjects who trained on individual cues directly. This finding also lends further credence to the view that a general mechanism limits integration capacity in human learners, rather than different factors such as integration and cue learning being different between groups.

3.8 Cross-experiment results

3.8.1 General visualizations

For completion and comparison, in this section, general visualizations are provided for each experiment. Psychometric functions (Eq. a.1) displaying how subject choice depended on integrated probability are provided in Figure 3-4. Learning curves for each experiment are visualized in Figure 3-5. Outcomes of the logistic regression analysis performed on cue weights in each experiment are provided in Figure 3-6.

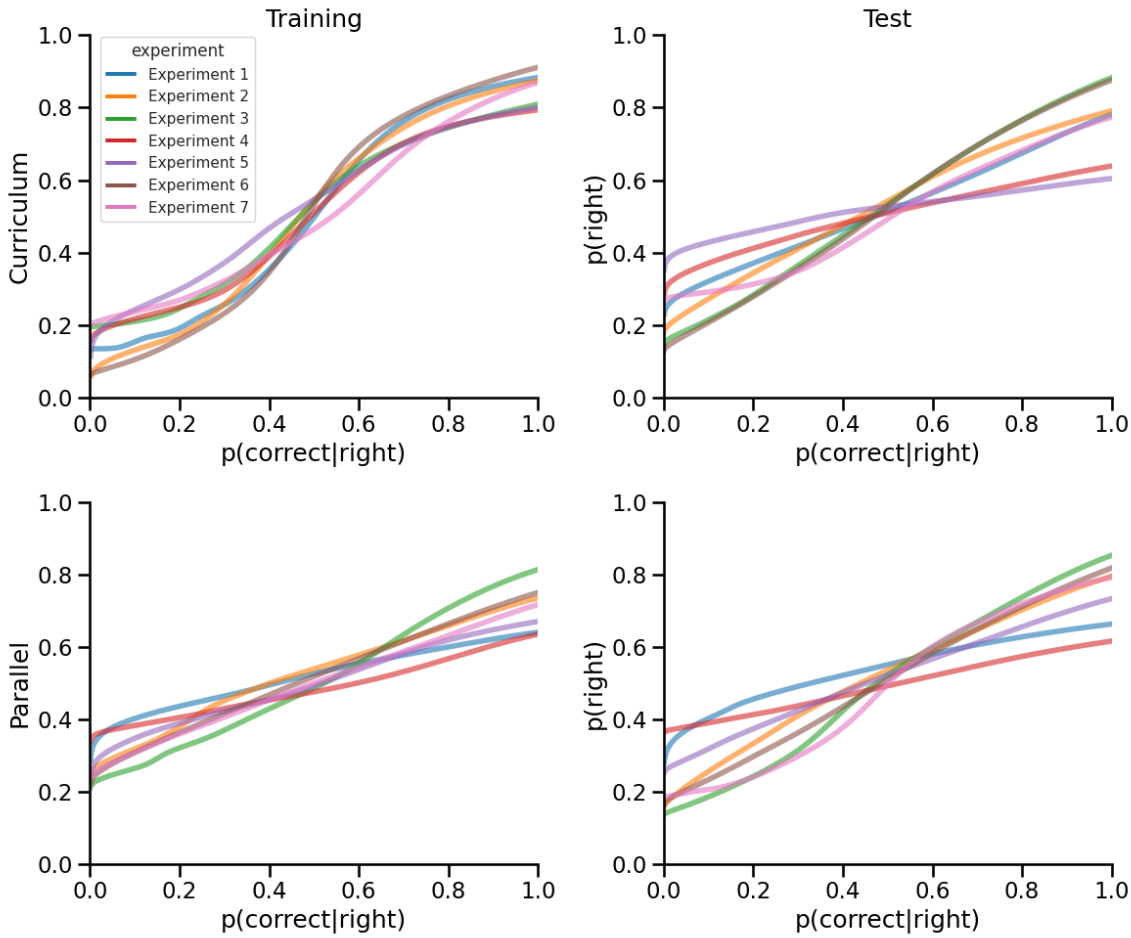

Figure 3-4: Psychometric function fits. Plotted is empirical probabilityof the rightward response as a function of the integrated probability of the rightward response being correct. Each line shows the average fit of a 4-parameter sigmoid function fitted to individual subjects, separately to the training and test phases. (**Top**) Curriculum condition. (**Bottom**) Parallel condition. (**Left**) Training phase. (**Right**) Test phase.

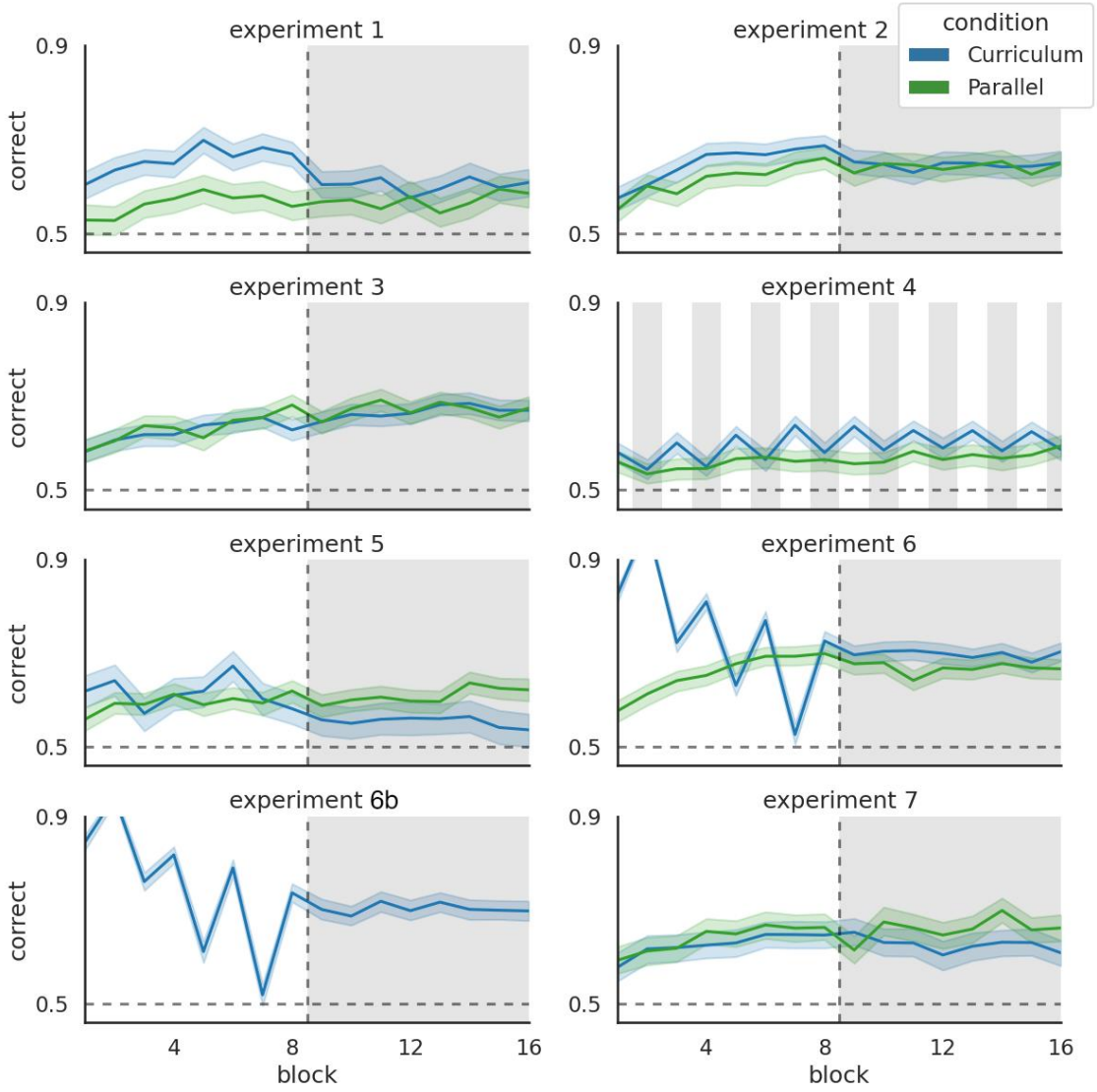

Figure 3-5: Learning curves in all experiments. Plotted is mean accuracy as a function of experimental block. Curriculum subjects are displayed in blue, while the parallel group is displayed in green. Horizontal dashed line indicates chance level. Shaded regions indicate the test phase.

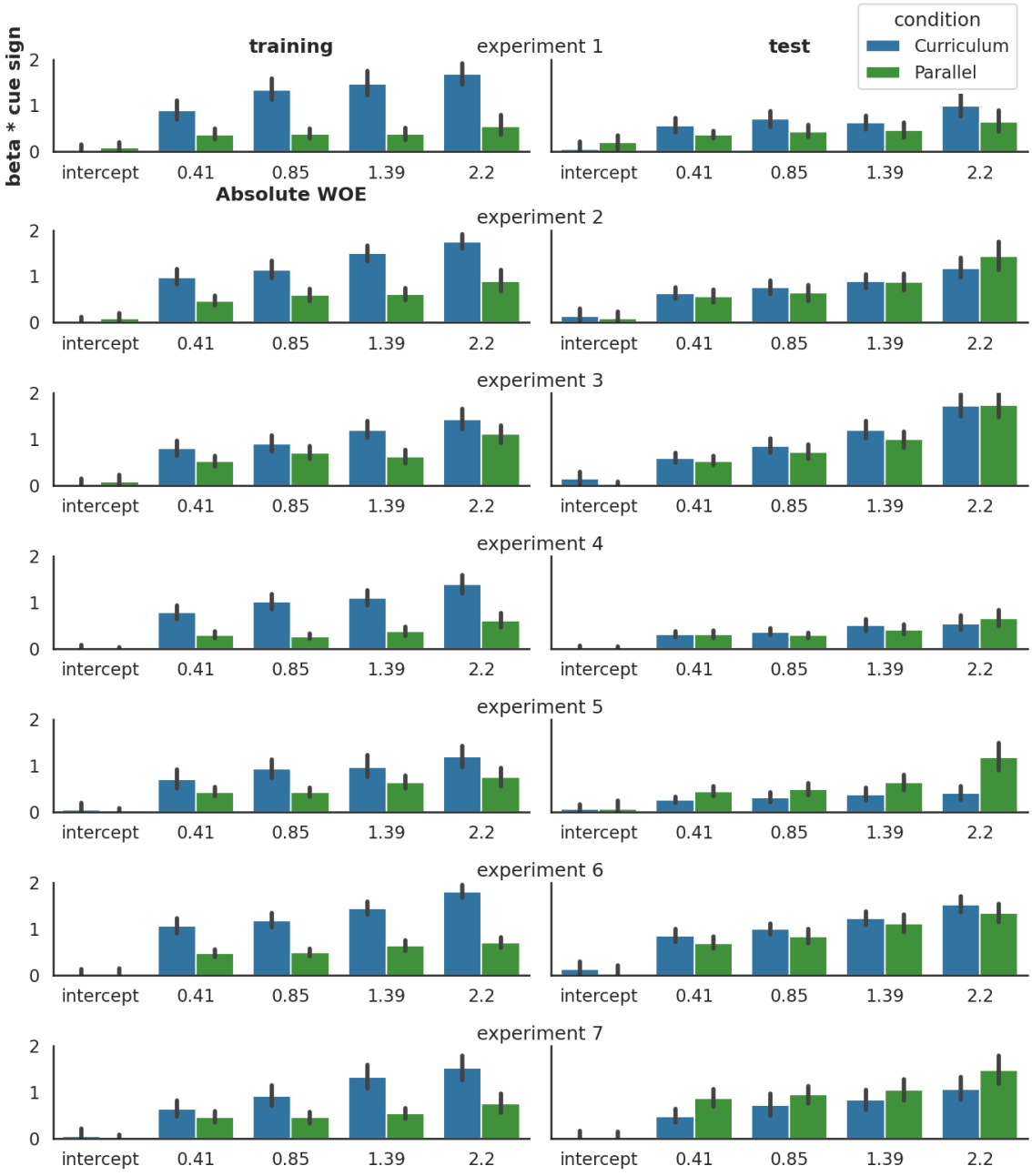

Figure 3-6: Learning curves in all experiments. Plotted is mean accuracy as a function of experimental block. Curriculum subjects are displayed in blue, while the parallel group is displayed in green. Horizontal dashed line indicates chance level, while the vertical dashed line indicates the transition from training phase to test phase.

3.8.2 No learning of integration rule during test

Full feedback was provided throughout the integration trials of the test phase in all experiments. This is particularly pertinent for the curriculum group, who in many experiments were trained exclusively on singletons, and so did not receive any information regarding the correct integration rule prior to test. A relevant question is whether these participants learn during the test stage, indicating separable learning of cue weights and integration rule. To determine this, the best linear fit of test trial accuracy (in experiment 7, only on integration test trials) was determined per-subject as a function of test trial number (1 for the first test trial, 2 for the second, etc.). The best-fitting intercept and slope were determined using the normal equation. If learning systematically occurred during during the test phase, this should correspond to test slopes that are consistently above 0 for different subjects.

Table 3-3 displays the outcomes of per experiment, per-condition 1-sample T-tests against 0. Figure 3-7 displays per-subject test slopes. Test slopes were not systematically different from 0 in any of the conditions where subjects were exclusively trained on singleton trials (the curriculum conditions of experiments 1, 2, 3 and 7). The only experiment in which participants showed evidence of learning in both conditions during test was experiment 4, where training and test blocks were alternated. Test slope difference from 0 also reached significance in the parallel condition in experiment 5. As this condition was similar or identical to the other parallel conditions, this result is likely a type I error.

Experiment	**1**	**2**	**3**	**4**	**5**	**6**	**7**
T statistic (Curriculum)	0.097	-0.184	1.951	2.523	-0.787	-0.650	-1.823
p-value (Curriculum)	0.923	0.855	0.060	**0.016**	0.443	0.519	0.958
T statistic (Parallel)	0.953	0.281	0.651	2.451	2.296	0.033	1.540
p-value (Parallel)	0.352	0.780	0.520	**0.019**	**0.028**	0.973	0.138

Table 3-3: Test slopes per experiment per condition. Displayed are *T*-statistics and corresponding *p*-values of 1-sample T-tests against 0.

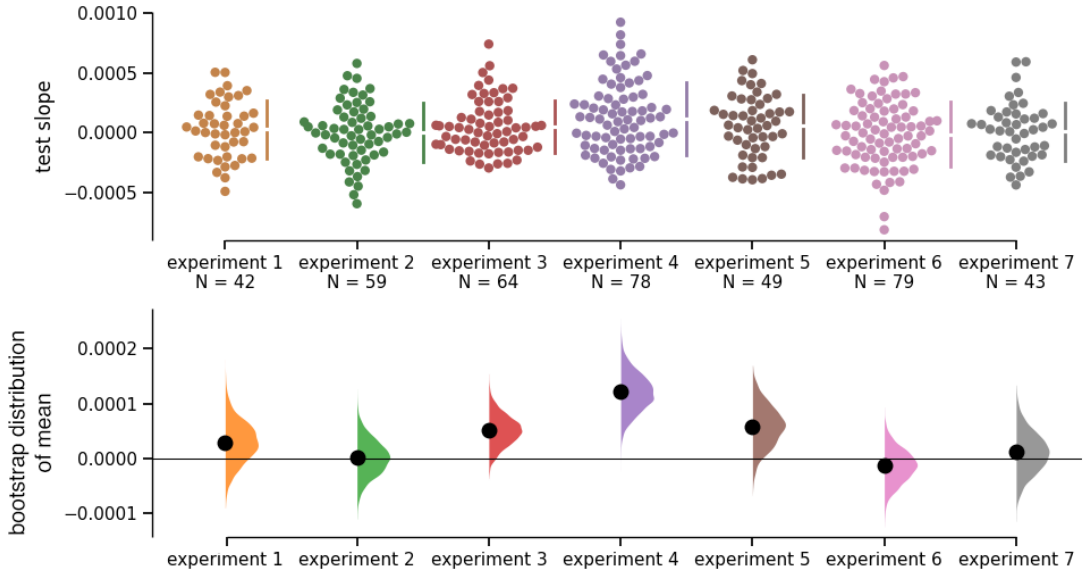

Figure 3-7: Linear model slopes for test phase. (**Top**) per-experiment per-subject test slopes of a linear model, predicting accuracy on integration trials in the test phase as a function of the number of test trials that had been observed. (**Bottom**) bootstrapped distribution of the mean in each experiment, generated over 5000 iterations. Black line indicates 0.

3.9 Modelling

Why do participants trained on single-cue curricula display learning benefits during training, but not during a multi-cue test? Relatedly, why do participants trained on multi-cue training schedules perform better on singleton trials than on multi-cue trials? As appropriate cue weights were obtained regardless of training curriculum, we reason that the bottleneck must be in the integration of information from different cues. The current section will describe a modelling analysis aimed at testing whether the data are well-described by integration based on heuristics, rather than using the ground truth integration rule. Two heuristics are examined as candidate mechanisms. The first option we consider is that subjects use a tallying heuristic, whereby the only information they use about stimuli is whether they are positive or negative, and decisions are based on the sum over present cues of these binarized quantities. In that case, performance should be preserved on singleton trials, but not on parallel trials, where cue validity cannot be ignored without incurring an accuracy cost. An alternative heuristic strategy that we consider is *k*-take-the-best, in which subjects consider only the *k* cues with the highest

validity. Under this model, a greater proportion of information is discarded for parallel trials than for singleton trials. Both heuristic strategies offer competing explanations for why performance on parallel trials is impaired relative to singleton trials, but differ in their specific trial-to-trial predictions.

3.9.1 Methods

Base model

We use the model described in Kurzawa et al. (2017) as a baseline. This model casts the WPT as a reinforcement learning problem, where the value y_i of either response option *i* is determined as the sum over cues *j* of the products between weights q_{ij} and their respective cue counts x_j (Eq. 3.6). We index cues in ascending order of probability, so that e.g. x_1=2 would correspond to two copies of the .1 cue being present in a trial.

$$y_i = \sum_{j=0}^{8} q_{ij} x_j \quad \text{(Eq. 3.6)}$$

Based on these per-action values, the response option is determined using a softmax function with inverse temperature ß, where ß is a hyperparameter that controls the amount of choice stochasticity. The resulting *Pi* is the probability of selecting action *i:*

$$P_i = \frac{e^{\beta y_i}}{\sum_{u=1}^{2} e^{\beta y_u}} \quad \text{(Eq. 3.7)}$$

After this, the obtained reward r_i was 1 for the rewarded action, and 0 for the unrewarded action. Cue weights are initialized as $\forall_i \forall_j$ q_{ij}=0. Cue weights were updated on every trial using a delta rule with learning rate α:

$$q_{ij} = q_{ij} + \alpha(r_i - y_i)x_j \quad \text{(Eq. 3.8)}$$

We note that equations 3.6-3.8 are adapted from Kurzawa et al. (2017) and do not reflect our own contributions. We make one general modification. Whereas the original model only updates the weights for the chosen action, we also perform the counterfactual update. In other words, we assume that subjects understand that if one response option was correct, the other was incorrect, and vice-versa.

This simplifies the learning dynamics, which then becomes independent of softmax-based choice sampling. Furthermore, given an empirically observed schedule of per-trial cues and reward assignments, the learning dynamics of the base model then becomes deterministic given α and ß.

Tallying model

The base model described so far performs lossless integration of cue weights on all trials. In the tallying model, the value function (Eq. 3.6) was modified with the addition of a signum function:

$$y_i = \sum_{j=0}^{8} sgn(q_{ij}) x_j \qquad \text{(Eq. 3.9)}$$

$$with\ sgn(x) := \begin{cases} -1\ if\ x < 0, \\ 0\ if\ x = 0, \\ 1\ if\ x > 0. \end{cases}$$

This manipulation was applied for action selection (Eq. 3.7), but not for weight updating (Eq. 3.8), so that the weight learning dynamics remained unaltered. This being the case, cue weights q_{ij} in the tallying model may be interpreted as a measure of certainty of a cue's predictive direction, as cue weights with greater absolute values require more contradictory evidence in order to change sign.

k-take-the-best model

In the k-take-thebest model, all cues present on a given trial are first sorted in descending order of their validity $|q_{is}|$, indexing now over the ordered set of present cues with *s* instead of over all cues with *j*. Again a modified value function was used (Eq. 3.10), this time to use only a subset of cues, which depended on the hyperparameter *k*, $k \in \mathbb{R}$ with $0<k<4$. We use *x mod y* to denote the modulo operation. For non-integer choices of *k*, the final cue was only partially used in the integration. For example, for $k=3.4$, the term $k\ mod\ 1$ returns 0.4, so that only 40% of cue $q_{i,4}$ would be used.

$$y_i = \sum_{s=1}^{\min(Ncues,\ floor(k))} (q_{is}) + (k\ mod\ 1) \cdot q_{i,ceil(k)} \qquad \text{(Eq 3.10)}$$

Our formalization of TTB is flexible, encompassing the original, single-reason decision-making version of TTB (k=1), while being capable of approximating the base model, which uses a fully compensatory strategy (k=4), along with any interpolated model in-between.

The TTB heuristic is normally used in the decision making literature, where it is commonly assumed that all cue weights are known a priori (Czerlinski et al., 1999; Gigerenzer & Goldstein, 1996; Parpart, Jones, & Love, 2018). As a consequence, it is not well-established whether this process should affect only decision-making, or also learning. In all analyses, we include a model version which used Equation 3.10 for action selection, but equation 3.6 for weight updating (TTB-all, so named for using all present cues in its weight updates), and also a version in which Equation 3.10 is used for both action selection and weight updating (TTB-used, in which only used cues are updated). An overview of the model space, including fitted hyperparameters and which value function is used in decision making and weight updating in each model is provided in Table 3-4.

	Base model	*Tallying*	*TTB-all*	*TTB-used*
Hyperparameters	α, β	α, β	α, β, k	α, β, k
Decision-making (Eq. 3.7) y_i used	Standard (Eq. 3.6)	Binarized (Eq. 3.9)	Most predictive k cues only (Eq. 3.10)	Most predictive k cues only (Eq. 3.10)
Update Rule (Eq. 3.8) y_i used	Standard (Eq. 3.6)	Standard (Eq. 3.6)	Standard (Eq. 3.6)	Most predictive k cues only (Eq. 3.10)

Table 3-4: Model space overview. Displayed are the models used, their hyperparameters, and which value function is used for decision making and weight updating, respectively.

Model fitting

The above models were fitted to the per-trial responses using the negative log-likelihood (NLL) of subject test responses under the model as the optimization objective. Hyperparameters were fitted per model per subject using a global optimization algorithm (dual annealing from Python's SciPy package), with parameter boundaries $\alpha \in (0, .2)$, $\beta \in (0, 10)$, $k \in (0, 4)$. Models were trained using the exact

trial structure and reward assignments that real subjects faced. BIC (Bayesian information criterion) was then computed for the best-fitting versions of each model for per subject. We opted to use BIC as a comparison metric as our TTB models introduce an additional parameter (k), and we wanted a strong test to ensure that this additional complexity was warranted. The BIC penalizes complexity more strongly than raw NLLs, and when more than 7 datapoints are fitted per subject , also than its common alternative, the Akaike information criterion, (Vrieze, 2012).

3.9.2 Results

Per-experiment BIC fits are displayed in Figure 3-8. In all experiments, the TTB-used model explains the cohort-level data best, having the lowest BIC. The same result is obtained within all parallel groups. In contrast, the tallying model had lower BIC in the curriculum condition than in the parallel condition, with the exception of experiments 3 and 4 (which used 2 cues for integration trials and interleaved training and test blocks, respectively). Of further note is that neither the base model, not the TTB-all model provided the best explanation in any experiment, or any condition within experiments.

To gauge whether these models are fitting accurately, Figure 3-9a displays the negative log-likelihood (NLL) of the TTB-used model minus the NLL of the base model in experiment 6. The TTB-used model obtains lower NLL for subjects whose best-fitting k is substantially lower than 4, and approximately identical likelihoods when k approaches 4. Figure 3-9b displays the empirical differential between singleton trial test accuracy and parallel trial test accuracy in experiment 6. Here, subjects best fitted by any of the heuristic models had higher accuracy on singleton test trials than on parallel test trials, Wilcoxon signed-rank test $T \geq 138$, $p \leq 0.016$, while this is not true for subjects best fitted by the base model, $T = 269$, $p = .682$. This provides a correspondence between model and subject behavior, as the heuristic models were employed to explain differences between singleton and parallel trial performance.

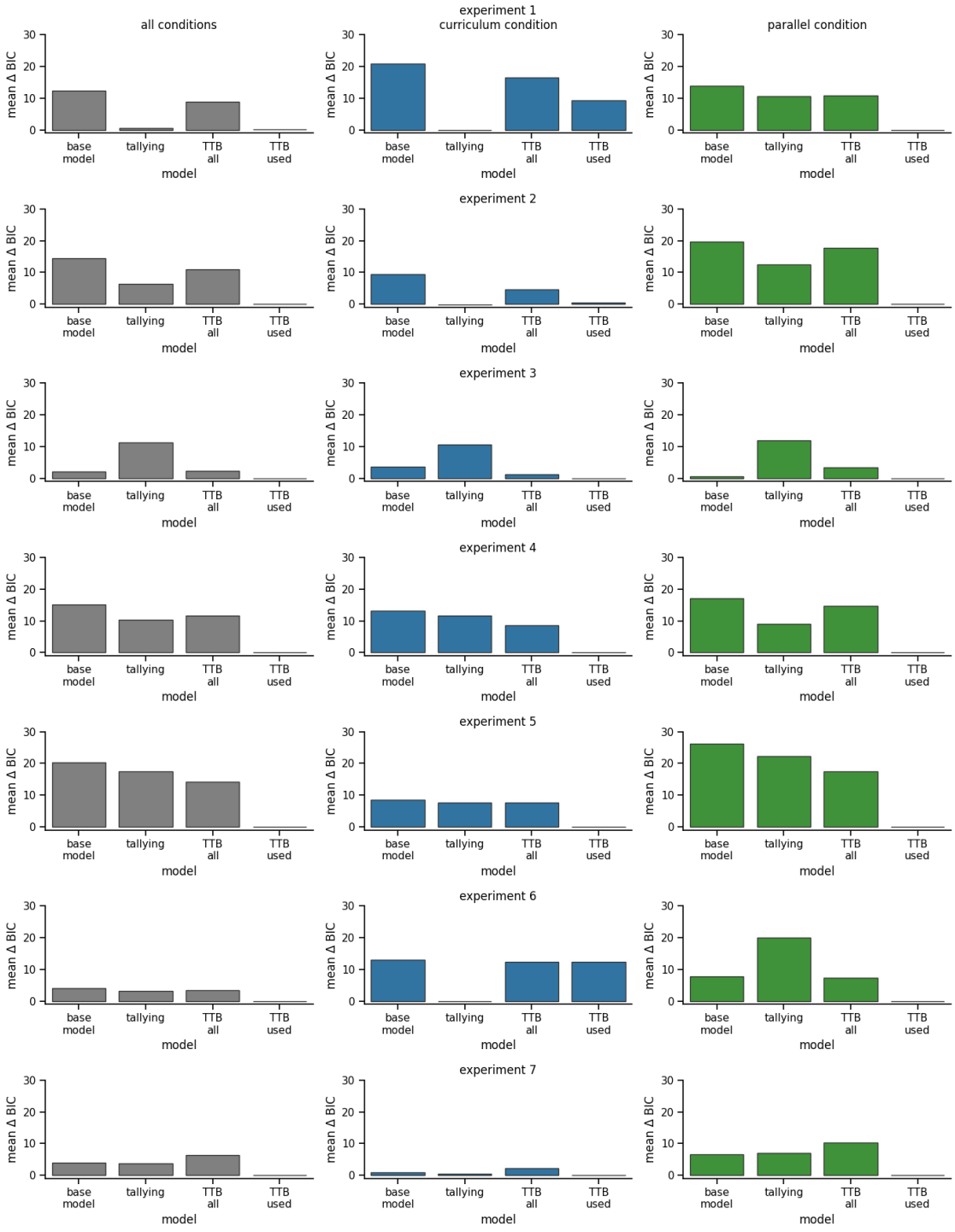

Figure 3-8: Model fitting results of heuristic-based models. mean BIC difference relative to best-fitting model. Rows display results for the different experiments. (**Left**) cohort-level (pooled over conditions for each experiment). (**Middle**) curriculum condition. (**Right**) parallel condition.

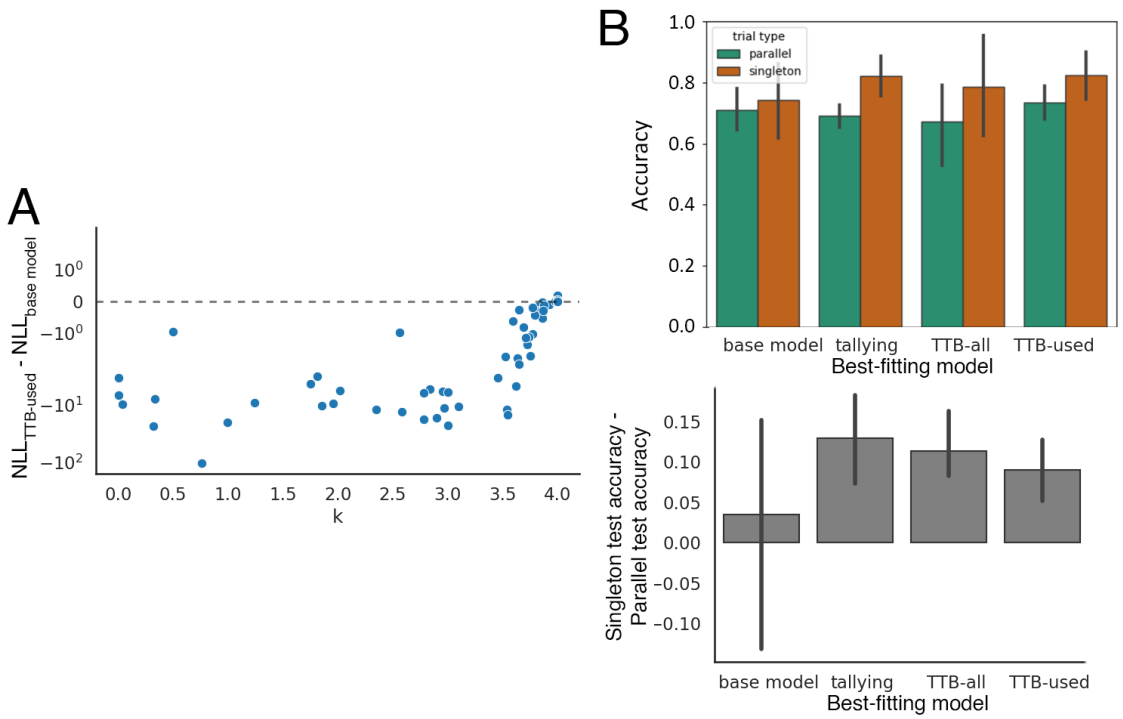

Figure 3-9: Qualitative features of the fitted models. (**A**): Relationship between NLL(TTB-used) – NLL(base model) and k of TTB-used model in experiment 6. (**B**): Difference between singleton test trial accuracy and parallel test trial accuracy for subjects best fitted by each model in experiment 6. (**Top**): absolute accuracy for parallel and singleton test trials. (**Bottom**): per-subject differential in test accuracy between singleton and parallel test trials.

3.10 Chapter discussion

In this series of experiments, our original intent was to investigate whether an appropriate training schedule can promote learning beyond what is learned through training on multi-cue examples in the probabilistic information integration learning setting. We reasoned that as multi-cue examples involve a complex credit assignment problem, subjects may benefit from learning the cues in isolation, with positive transfer to the multi-cue setting. Across 7 experiments, we do not find consistent evidence for this claim. Instead, subjects trained on multi-cue trials were unimpaired in singleton trial performance and per-cue estimates relative to subjects trained on single cues, suggesting that instead, human learners are remarkably adept at solving the credit assignment problem.

In addition, a novel insight acquired from these experiments is that human probabilistic information integration appears to be subject to a limit in integration capacity. The combination of strong credit

assignment paired with limited ability to integrate is demonstrated most clearly in experiment 7. In this experiment, subjects trained only on parallel trials performed better on singleton cues than on parallel trials. Moreover, they were as good at providing single-cue estimates as subjects trained only on single cues. This is surprising, as the mean WOE is larger when multiple cues are presented. Indeed, the base model from the current literature (Kurzawa et al., 2017) predicts that subjects should do better on singletons than on parallel trials, the converse of what we observed.

The finding that regardless of curriculum, subjects performed worse on trials with multiple cues underscores the importance of understanding limitations in human information integration ability. As a step in this direction, we examined how well our data was explained by two heuristic decision strategies: tallying and take-the-best. In this regard, we find that the TTB heuristic provides the best explanation of the data in all evaluated experiments. This model consistently provided better fits than the base model and tallying model, despite application of a Bayesian complexity penalty to account for its higher complexity. However, the decision strategy that is used may depend on the training conditions, as in test responses of subjects in the curriculum condition were relatively well-fit by the tallying model, compared to test responses of subjects in the parallel conditions. Interpreting this result, previous work has shown that use of a tallying strategy is promoted when the direction of cues is known, but their ranking is not (Binz et al., 2020). We offer that training on singleton trials may have emphasized cue signs, as cue validity was irrelevant during training. Similarly, in experiment 6, introducing the cues in equal-validity pairs may have fostered a cue sign-based representation, and consequently a tallying strategy. Our finding that subjects use a limited subset of cues relates to a previous finding where subjects were found to base decisions primarily on a single attended cue in a 2-cue setting (Akaishi et al., 2016). There, shifts in which cue was attended occurred when outcomes were inconsistent with subjects' single-cue based predictions. However, Akaishi et al. used a design where cue values changed over time, irrelevant cues were used, and relevant cues could later become irrelevant, providing several conditions which may have encouraged a single-cue strategy. The finding that the data is overall best explained specifically by the version of TTB where discarded cues are not updated also suggests a mechanism of how humans may overcome the credit assignment problem. Applied to the example from the introduction, the combinatorial problem of predicting the weather given

many cues may be complex, but if we wish to learn if weather forecasts are reliable, we can simply ignore all other cues and test how often the forecast was right. While this method ignores conditional probabilities and so may not guarantee optimal learning, robustness to correlations between cues is precisely the property of single-reason decision making to which its success is usually attributed (Gigerenzer & Todd, 1999; Parpart et al., 2018), so that applying this heuristic both during learning and inference is an internally consistent strategy.

The TTB heuristic originates from the decision-making literature, and our work is novel in bringing this heuristic to the learning setting. As a result, it has not been established how TTB learners, characterized by ignoring part of the available information, learn the cue validities that their decisions rely upon. In the decision-making literature, this problem is commonly circumvented by assuming that the learner knows all cue validities beforehand (Gigerenzer & Goldstein, 1996; Parpart et al., 2018). In our modelling, we have compared a version of TTB in which learning of cue validities was unimpaired (TTB-all) with a version in which only the cues used for decision-making were updated (TTB-used). Across all examined experiments, TTB-used consistently provided a better explanation of the data. More broadly, this suggests that TTB may have an important role in learning, not only in decision making. This is of interest particularly because this heuristic can have extreme effects on information sampling, so that the ideal per-cue estimates assumed in most decision-making studies may never be obtained. An interesting direction for future research would be to test this more directly, creating settings in which one can accurately trace how early information constrains subsequent sampling, potentially forming “learning traps” (Rich & Gureckis, 2018) even if full feedback is provided.

Chapter 4: Compositional rule learning

4.1 Chapter Abstract

Several strands of contemporary research in neuroscience and artificial intelligence are inspired by humans' proficiency at adapting to novel situations, attributed in part to their ability to leverage compositional structure. However, the precise behavioral phenomenon – when humans do and do not succeed at compositional generalization – has not been fully characterized. In the present work, we sought to determine what factors determine compositional generalization success in healthy human adults. We operationalized this question using a task with compositional stimuli, consisting of a color and shape. These corresponded to locations on a 2D map, which participants were asked to predict. Each input dimension corresponded to a distinct rule. To predict the outcome locations for held-out color-shape combinations, participants had to correctly infer and combine the underlying rules. Across four experiments with a total of 605 participants, we find consistent evidence that generalization success is promoted by curricula that emphasize the underlying structure: blocked training presentations, and alignment between rule and stimulus axes. In addition, a substantial number of responses were 'unilateral' – incorrect in one dimension, but consistently correct in the other. Finally, participants exhibited compositional generalization, applying linear rules in the linear setting, and nonlinear rules in the nonlinear setting. By contrast, standard feedforward neural networks produced linear generalization patterns, even when the underlying rules were nonlinear. We propose that participants´ behavior is best explained by a mechanism that separates representation of meaningful clusters, such as individual rules, from a controller that combines these. Simulations with this setup produced unilateral errors, benefitted from blocked training and exhibited compositional generalization of nonlinear rules.

4.2 Introduction

It is often asserted that humans are good at adapting to novelty (Passingham & Wise, 2012; Penn et al., 2008). For evidence of this versatility, one only has to consider the quotidian activities that most people master as adults. Navigating the local environment, managing household resources and interacting with social others all frequently require flexible responses to unanticipated challenges. Indeed, success at handling novelty has been touted as the key ingredient in human intelligence (Behrens et al., 2018; Ferguson, 1956; Lake et al., 2017), and is frequently counterpointed with the narrowness and rigidity of current AI systems (Marcus & Davis, 2019).

Whilst anecdotal evidence of human mental versatility abounds, capturing this phenomenon in the lab has proved remarkably challenging. People can of course readily perform complex tasks when offered clear instruction in natural language. However, it is less obvious whether they spontaneously solve new puzzles without verbal instruction. Dealing with unfamiliar situations requires the repurposing of existing knowledge and skills to novel settings, which is called generalization or transfer. The scope and limits of human transfer were the focus of a voluminous literature in cognitive psychology towards the end of the 20th century. Many studies asked whether reasoning puzzles (such as how to direct X-rays to destroy a tumor) were easier to solve after encountering problems with an analogous solution (such as how to deploy an army to conquer a fortress). Prior exposure to test items with common structure sometimes improves performance, but appeals to transfer have proved hard to disentangle from more mundane explanations that rely on hints or implicit instructions from the researcher (Barnett & Ceci, 2002; Gick & Holyoak, 1980). Moreover, human transfer abilities seemed to depend on transfer distance – generalization is often successful when old and new problems share physical features (near transfer) but frequently fails when problems share common structure but are superficially distinct (far transfer). This graded dependence on transfer distance makes it hard to make concrete claims about whether humans are adept at dealing with novelty or not, and has ultimately led to pessimistic claims that far transfer – moments of insight that link entirely distinct problems – are vanishingly rare (Detterman, 1993).

Part of the reason why the detailed study of human transfer yielded unsatisfying conclusions is that researchers of the 1970s and 1980s lacked a computational language for quantifying their study of transfer learning. A relevant framework has arisen over recent years as transfer learning has become a core theme in machine learning. Machine learning researchers conceive of learning as an optimization process in which the parameters of a function are adjusted until inputs are mapped onto desirable outputs. In supervised learning, successful generalization requires new and unseen inputs to be mapped to the ground truth defined by an external oracle or teacher. Near transfer is routinely the reported success metric when training neural networks. For example, after being trained to label objects, researchers measure network accuracy on new, unseen instances of the same object classes (Russakovsky et al., 2015). Near transfer can be achieved if the function mapping acquired during training is sufficiently smooth to allow interpolation or extrapolation to similar but nonidentical training exemplars, and a large literature explores the principles that allow this to occur (Ahmad & Tesauro, 1988; Belkin, Hsu, Ma, & Mandal, 2019; Belkin, Hsu, & Xu, 2020; Novak, Bahri, Abolafia, Pennington, & Sohl-Dickstein, 2018; Zhang, Bengio, Hardt, Recht, & Vinyals, 2021).

Far transfer, however, remains an elusive goal in AI research. One promising solution revives an old idea in cognitive science – that thought and action are fundamentally compositional. The hypothesis is that the world is structured so that new tasks can often be solved by combining old ones, and thus systems with an inductive bias to explicitly compose new knowledge and skills from existing building blocks may succeed in novel settings (Hupkes, Dankers, Mul, & Bruni, 2020). For example, an agent that has learned to grasp and walk might be prone to combine them to transport an object from one setting to another. Composing behaviors in this way has the merit of limitless generativity, which was a fundamental argument for the compositionality of language put forward by early cognitive scientists (Chomsky, 2017). However, canonical neural network models do not naturally exhibit compositional behavior (Hupkes et al., 2020; Lake & Baroni, 2018), and instead require architectural innovation (Chang et al., 2018) or the addition of neuro-symbolic features (Nye, Solar-Lezama, Tenenbaum, & Lake, 2020) to solve far transfer tasks.

How do canonical neural networks respond to compositional structure? We propose that neural networks have a "linearity bias", whereby they have a tendency to represent problems as linear transformations when the induction problem is sufficiently underconstrained. This is true by default in their most basic form — linear neural networks — as these are fundamentally incapable of representing nonlinear transformations. In practice, nonlinear activation functions are commonly added to neural networks to increase their expressivity. Yet even in several popular types of nonlinear neural network, the early dynamics are dominated by the linear solution. To see why this is the case, consider that networks are typically initialized with small weights (Glorot & Bengio, 2010; He, Zhang, Ren, & Sun, 2015). In this zero-proximal regime, the hyperbolic tangent approximates the identity function, so that the early dynamics resemble those of a linear neural network. A similar argument can be applied to the sigmoid function, which approximates an affine function in this regime. If the linear solution obtained through these early dynamics fully explains the training data (which is particularly likely when the learner must generalize from few examples), learning stops there, and generalization will obey the linear solution. If not, learning will proceed to the nonlinear parts of these activation functions, gradually modifying the linear solution to account for nonlinear structure.

On the whole, the determinants of human generalization success and failure remain unclear, and we know little about the cognitive and computational mechanisms that subserve it (with a few exceptions, such as compound generalization, Soto, Gershman, & Niv, 2014, and fear generalization, Dunsmoor & Paz, 2015; Dymond, Dunsmoor, Vervliet, Roche, & Hermans, 2015). In the current chapter, we set out to study the limits of human transfer, and the factors that promote or impair compositional generalization. Our first contribution is to define a supervised task that neatly exposes the computational challenge of transfer learning, because it comprises two problems with identical functional form, one which requires explicit composition (as defined above) and the other that does not. We predicted that humans, but not vanilla neural networks, would succeed at the explicitly compositional task. Our hypothesis is that human learning involves solving tasks through inferring their structure. As a consequence, we expected that subjects would benefit from conditions which clarified this structure, including (i) chunking rules in time and (ii) training on cleanly dissociable task factors. In

addition, a predicted hallmark of compositional learning is that per-rule learning should display a greater degree of asynchronicity than would be predicted by an unstructured account.

4.3 Methods

4.3.1 Participants

In total, 605 subjects (293 female) participated in the studies described here. Participants were recruited on the crowdsourcing platform Prolific and were rewarded £9, plus a performance-based bonus of up to £6. An age range restriction of 18-40 years was applied and participants were required to have a submission approval rate of at least 85% over at least 5 prior submissions. The data rejection procedure is described in section 4.3.5. All experiments were approved by the Medical Sciences Research Ethics Committee of the University of Oxford (approval reference: R50750/RE001).

4.3.2 Design

In this series of experiments, our main aim was to probe the mechanisms by which humans generalize when they encounter novel combinations of familiar stimulus dimensions. To this end, participants were trained to indicate the location associated with 25 stimuli, each composed of a color and an animal shape. 5 different colors and shapes were used, and stimuli were always presented on a square, dark grey background. Participants responded by dragging a pirate avatar to their estimate of the correct location on a circular map in a computerized task. Crucially, subjects received feedback only for 9 out of 25 combinations (feedback locations), featuring each level of each dimension at least once.

Each stimulus dimension indicated a separate rule, so that correctly inferring the underlying rules permitted appropriate generalization to those combinations for which feedback was never received (generalization locations). To test the generality and scope of our findings, two variations of the task were used, set apart by the type of rules they used. The “grid task” used a linear ruleset: one dimension (e.g. shape) indicated the horizontal position of the reward, and the other (e.g. color) its vertical position. The “polar task” instead used a non-linear ruleset, where one dimension indicated the angle, and the other the eccentricity of the reward location. These tasks are visualized in Figure 2. Mapping of stimulus dimension onto rule type and the ordinal mapping of dimension levels to rule

magnitudes were randomized between subjects. In the polar task, the ground truth locations were randomly rotated per subject, but are derotated into a common space in our visualizations.

4.3.3 Procedure

The trial sequence is displayed in Figure 4-1. Trials began with fixation, which was on-screen for 1000ms. After this, a large version of the stimulus was presented in the screen center for 1000ms. Then, this was replaced by the response space, with the pirate in a randomized starting location, and a smaller version of the stimulus displayed above the response space. The response space was a circle with a radius of 265 pixels, whose perimeter was indicated by a black line. A single dot was displayed in the center of the response space as a reference point. Participants indicated their prediction by dragging the pirate and pressing the ´d´ button. If this did not occur within 4 seconds, a visual timer appeared, which gradually became smaller. From onset, it took 20 seconds for this timer to fully deplete. At this point, no response was logged if the pirate was still in the starting position. If the pirate had been moved, its current position was taken as the participant's response. Upon timeout or response, the pirate disappeared, and a black dot appeared in its place, more precisely indicating the response location. On generalization trials, this lasted 1600ms. On feedback trials, this lasted 300ms, after which an "X" indicating the ground truth location was added to the display. After another 300ms, three concentric circles appeared around the X for 1000ms, indicating the maximum range at which points would be rewarded. On either trial type, the previous feedback then disappeared, and was replaced by the text "no feedback", "incorrect", or a treasure chest and the number of points earned on that trial, as appropriate. This feedback was displayed for 1300ms, after which the trial was concluded.

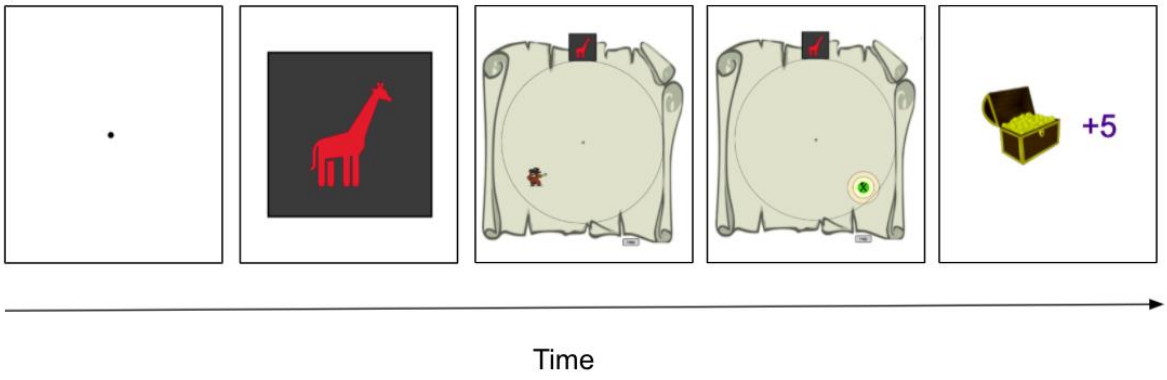

Figure 4-1: Example trial sequence for a feedback trial. All elements are to scale.

Points were awarded as follows: 5 points were awarded if the distance between response and ground truth location was less than 20 pixels. 2 points were awarded for a proximity of 20-40 pixels, 1 point was awarded for a proximity of 40-60 pixels, and 0 points otherwise. The 3 concentric circles displayed on feedback trials indicated the ranges at which 5, 2 or 1 points were awarded, and participants could see how many points they earned by which of the three circles was highlighted (or none in the case of no points). The number of points obtained during feedback trials in the current block was displayed on-screen. Participants were instructed that although they would not see how many points they earned during trials for which no feedback was provided, their performance on these trials would contribute toward their bonus payment. The bonus payout was the number of points earned in a block, with a rate of 1 penny per point, up to a maximum of 36 pence per block.

Each experiment consisted of 14 blocks. In each block, every stimulus (i.e. combination of color and shape) was queried exactly once. In the first 9 trials of each block, the feedback locations were queried, whose order obeyed a condition-specific curriculum, detailed in the next section. In the last 16 trials of each block, generalization locations were queried in randomized order. Between blocks, a pause screen was shown, displaying the bonus earned by the subject on the previous block and their accumulated total bonus. The participant could continue to the next block when ready by clicking the space button.

4.3.4 Experimental conditions

A total of 8 conditions were recorded, testing different determinants of learning success. The first set of experiments tested the effect of axis alignment. In the + curriculum, feedback was provided for the central two axes – the central row and column in the grid task, or the central ring and one spoke in the polar task. This meant that the set of feedback locations contained all shape levels for a single color, and all color levels for a single shape (e.g. all red stimuli and all crabs). As different stimuli within an axis involved changes in only one rule, we consider the stimulus dimensions and rules to be aligned in this setup. In the X curriculum, the stimulus axes were misaligned with the rule space, forming an X-shape in the grid task. The second set of experiments tested the effect of blocked vs. interleaved presentations. In these experiments, feedback was provided for a random row and column (grid task) or a random ring and spoke (polar task). Recall that the first 9 trials of each block query the feedback locations. In the blocked curriculum, trials 1-4 queried locations from one axis, trial 5 the location shared between training axes, and trials 6-9 the locations from the other axis. Which axis was queried first alternated between blocks, with the dimension that was queried first in block one randomized between subjects. In the interleaved curriculum, feedback locations were instead presented in randomized order. The +/X curricula were always blocked. We refer to +, X, blocked and interleaved as the curricula used in this study. A grid task cohort and polar task cohort were collected for each of these four curricula, for a total of 8 conditions. A visual overview of the experimental conditions is provided in Figure 4-2.

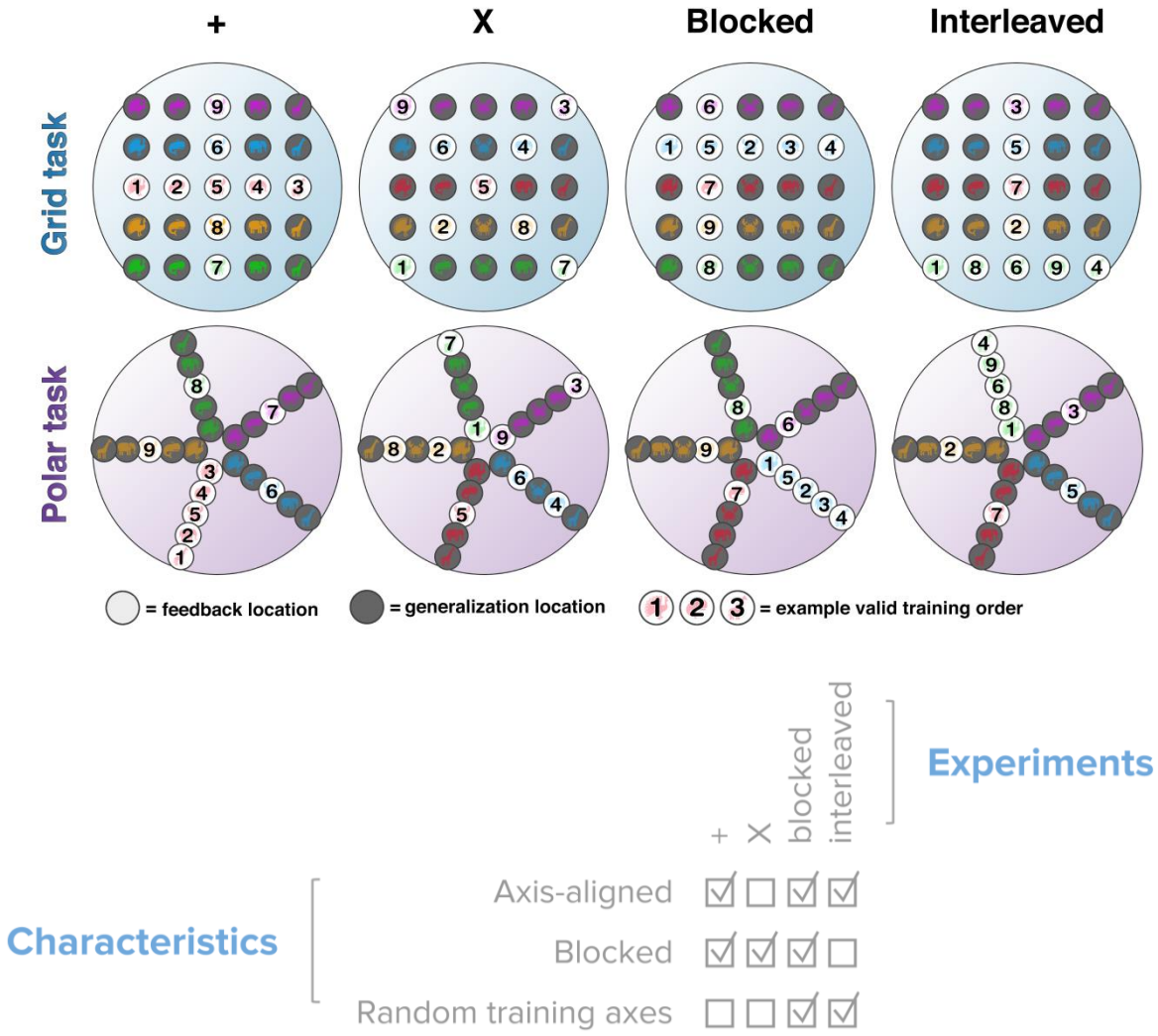

Figure 4-2: Overview of experimental conditions. 4 curricula were tested, each once with the linear ruleset and once with the non-linear ruleset, for a total of 8 conditions. In all conditions, subjects received feedback only on two axes in the stimulus space. The +/X experiments differed in whether participants' training axes were axis-aligned with the rule axes. The blocked/interleaved experiments differed in the order in which feedback locations were presented. In the blocked curriculum, successive feedback examples would cover one full axis first, before proceeding to the other. In the interleaved curriculum, feedback locations were presented in a randomized order. The plots display a valid set of feedback locations (white background) for each condition, while generalization locations are displayed with a dark background. A valid feedback location order is indicated with numbers for the feedback locations. The exact feedback location order varied between blocks, but was blocked by axis in all curricula except the interleaved curriculum. Stimulus contrast is lowered for visualization purposes.

4.3.5 Data rejection

Data rejection was based on feedback location accuracy in the second half of the experiment. This choice was made as generalization location responses form our dependent measure, and learning the feedback locations is a prerequisite for appropriate generalization. Responses were considered to be correct if they were within 60 pixels of the ground truth location. As in some conditions, memorizing feedback locations may have been more difficult than in others, our criterion was the within-condition median absolute deviation, using the default scale constant of b = 1.4826 (Leys, Ley, Klein, Bernard, & Licata, 2013), and a rejection threshold of 3 median absolute deviations, described as "very conservative" in the same source. Application of this criterion resulted in a total of 127 rejections. Per-condition rejections are visualized in Figure 4-3.

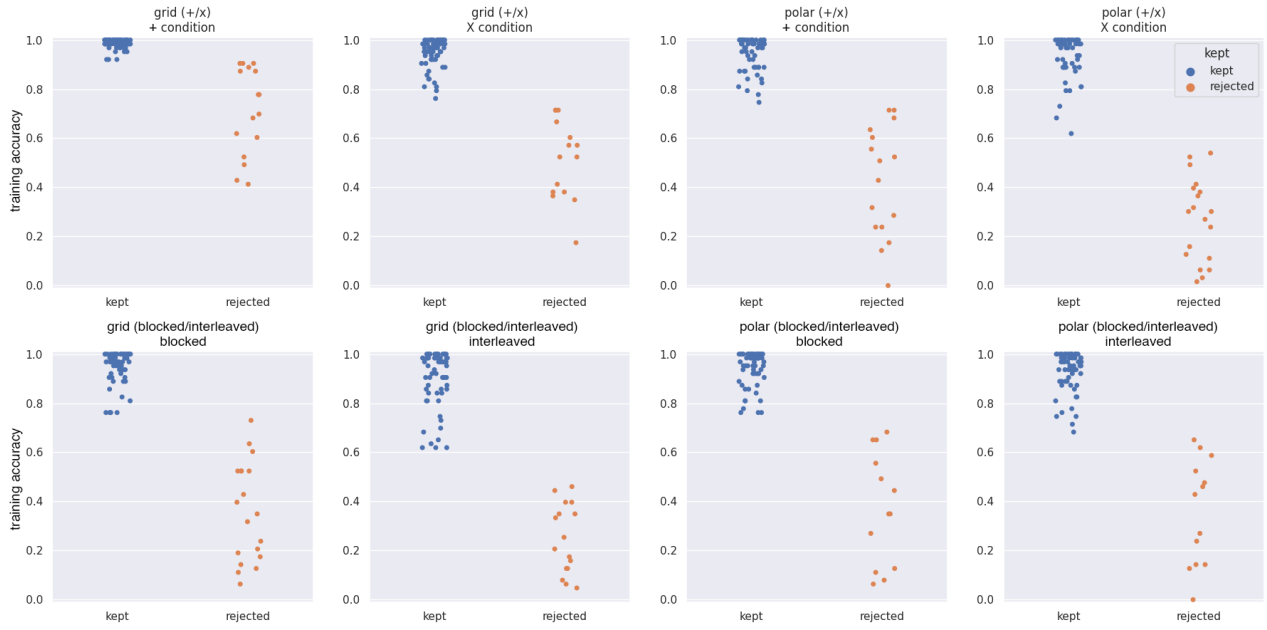

Figure 4-3: rejections per condition. Rejection took place within conditions, using a threshold of 3 median absolute deviations on mean training accuracy in the second half of the experiment. Displayed are second-half training accuracies for kept participants (blue) and rejected participants (orange) for each condition.

4.3.6 Stimulus dissimilarity rating task

The main experiment was preceded by a dissimilarity rating task, aimed at gauging participants' priors. In this phase, a set of stimuli using colors and shapes (5 levels each) not used in the main experiment

were used. Unlike stimuli from the main experiment, they had no background. The stimuli from the dissimilarity rating task appeared in random locations in the same circular response space that was used in the main experiment. Participants could drag and rearrange these stimuli. They were instructed to *"find a position for each object that someone else would find most intuitive, and easiest to learn"*. They could continue by clicking a *"Finish"* button, provided they had moved at least 50% of the stimuli. Clicking the finish button before this point resulted in the message *"You'll be able to continue after engaging with the task a bit more".* After completing the dissimilarity rating task, participants were advised that there was no connection between the task they had just completed and the main task. A bonus of £1 was awarded for completion of the dissimilarity rating task, with the remaining bonus of up to £5 being based on main task performance. Analyses pertaining to the dissimilarity ratings are described in section b.2.

Details on the neural network simulations and model fitting procedures are provided in the supplementary methods, contained in the appendix for this chapter.

4.4 Results

The induction problem in function learning is unconstrained – unobserved inputs may in principle correspond to any output. The type of generalization that is appropriate is consequently task-specific. Rather than taking a normative view, our initial focus is on the question by what mechanism human learners generalize in compositional settings and how this compares to generalization in canonical neural networks. In the introduction, we made the prediction that a canonical neural network would generalize according to a linear solution. To test this empirically, we first analytically computed a linear solution to our tasks, taking only the feedback locations into account (Figure 4-4, “Vector addition”, additional details in b.1.3). In the grid task, this linear solution succeeds at reproducing the ground truth generalization pattern, as the underlying rules are linear. In the polar task, the linear solution produces a generalization pattern that is distinct from the ground truth.

We next trained a canonical neural network containing a sigmoid nonlinearity on the task, predicting that its generalization would be well-characterized by the vector addition model. Mean location predictions queried at the end of 10.000 independent runs are visualized in Figure 4-4 (“Neural

network"). Indeed, the network's behavior qualitatively matches predictions from the vector addition model. Details of the modeling procedure are described in section b.1.4.

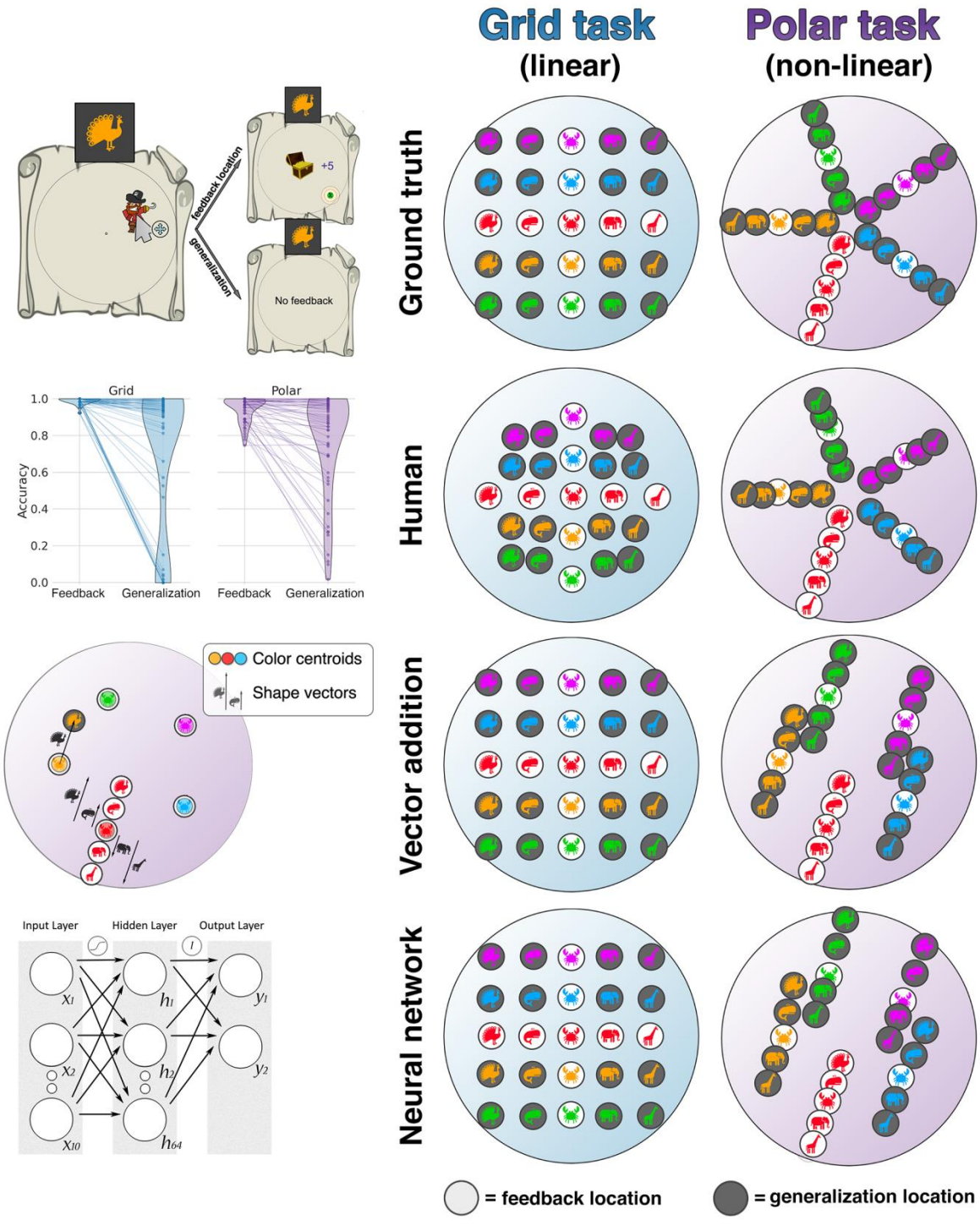

Figure 4-4: Compositional generalization. (**Left**) per-model visualizations. Top row: task setup. On every trial, the participant receives a bidimensional cue, on the basis of which they must predict the reward location by moving a pirate avatar. For feedback locations, the correct location is then shown, and points are awarded based on proximity. Second row: per-task accuracies for feedback and generalization locations in the last half of the experiment. Accuracy is defined as responding within a radius of 60 pixels of the reward location, matching the maximum distance at which points were awarded. Markers are individual subjects, violins are kernel density estimations. Third row: schematic illustration of vector addition model. This is a theoretical model which computes a linear transformation based on the feedback locations. Bottom row: neural network architecture. Details of the simulation setup are described in the supplementary methods. (**Right**) mean response locations for each model. Feedback locations are indicated by a white background, and generalization locations by a black background for visualization purposes.

Limiting ourselves to the + curriculum for comparison for the moment, we trained 153 subjects on these tasks (79 grid, 74 polar). Figure 4-4 ("Human") displays their accuracy scores and mean location predictions in the second half of the experiment. For this binarized accuracy measure, responses were regarded to be accurate when at least a single point was obtained. Overall, human learners qualitatively reproduced the ground truth pattern in both settings, applying linear rules in the grid task, and nonlinear rules in the polar task. This forms a contrast with generalization in the neural network model, whose generalization was approximately linear in both settings. However, as displayed in the violinplots in Figure 4-4, human generalization location performance featured substantial interindividual variability, with a subset of learners performing poorly on generalization locations despite high performance on feedback locations.

4.4.1 Determinants of compositional generalization success

In the basic + curriculum just considered, human generalization location performance was variable between participants, ranging from 0% to 100% accuracy. How do the training conditions predict generalization success? To address this question, we employed curricula aimed at understanding the effects of axis alignment (+ versus X curriculum) and the effects of blocking (blocked versus interleaved curriculum).

Specifically, we used a model fitting procedure, comparing the likelihood of a random model with the likelihood of a model describing appropriate generalization. We term the latter a "bilateral" model as it describes successful learning of both rules. In the bilateral model, the bulk of probability mass is concentrated on the ground truth, described by a narrow Gaussian distribution. In the random model,

every possible response location was equiprobable, modeled by the uniform distribution. Both models were evaluated on the generalization location responses of each subject. Additional details on the model fitting procedure are provided in the supplementary methods, section b.1.1. Per-subject log likelihood ratios (LLR) of these models are visualized for each experiment in Figure 4-5. In all experiments, generalization benefits were found in the predicted direction. The statistical results are provided in Table 4-1.

Task	**Curriculum**	**Median**	**Standard deviation**	***p*-value (bootstrap test)**
Grid	+	404	591	<.000†
	X	-647	532	
	Blocked	222	504	<.000†
	Interleaved	-76	581	
	All	87	588	<.000
Polar	All	321	432	
	+	380	419	.029†
	X	267	526	
	Blocked	413	504	.035†
	Interleaved	255	582	

†: one-sided

Table 4-1: Log-likelihood ratios of bilateral to random model per condition. Models were evaluated on all generalization location responses. Positive medians indicate higher relative fit of the bilateral model, while negative medians indicate higher relative fit of the random model. P-values indicate results of bootstrap tests computed on the difference in means between conditions, using 10.000 bootstrap iterations.

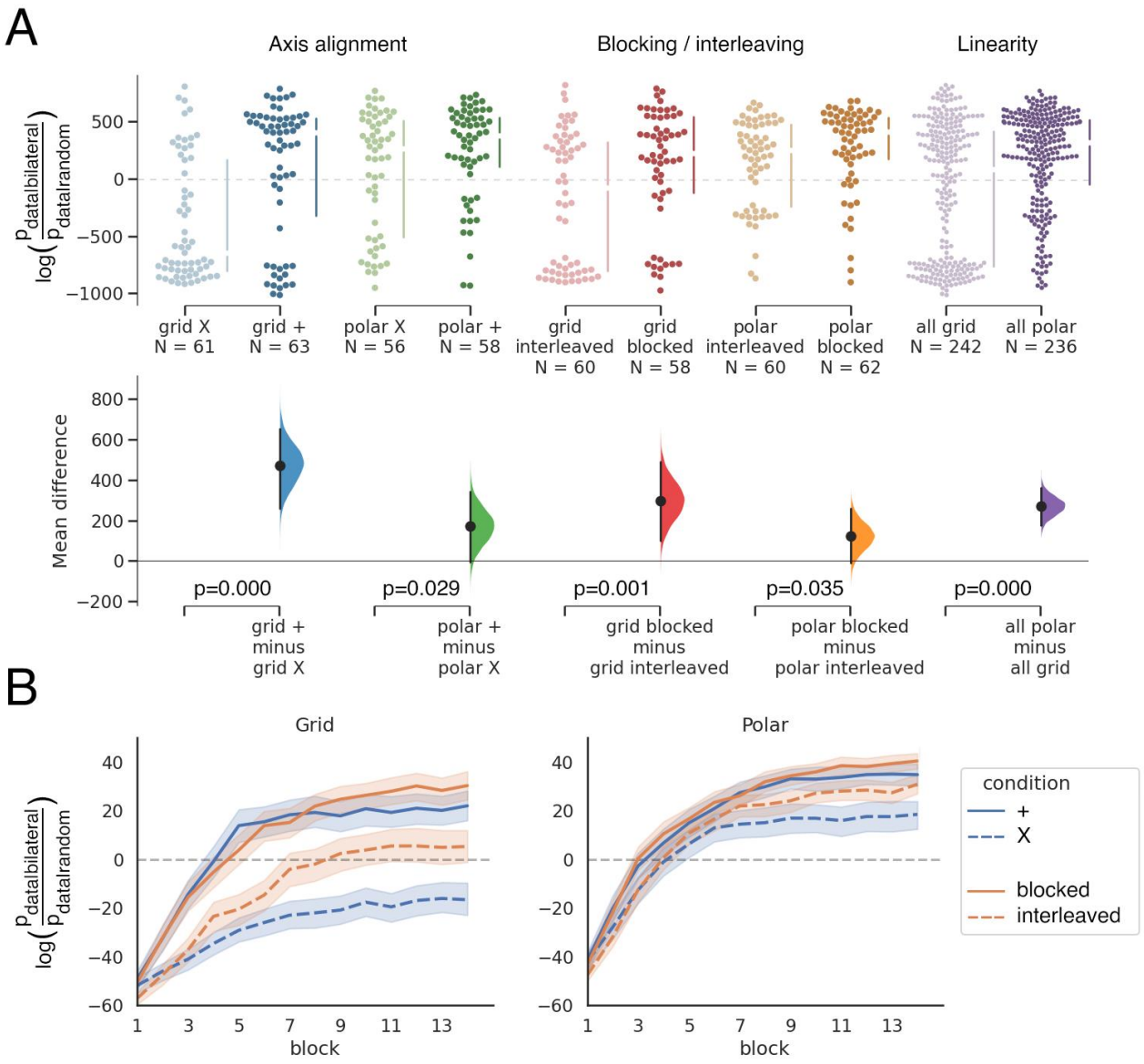

Figure 4-5: Effects of blocking, axis alignment and linearity on generalization success. (**A**) per-subject LLR between bilateral model (narrow Gaussian around ground truth) and random model (uniform distribution over all possible locations) for each condition. In addition, the rightmost plot displays LLRs pooled over all subjects who performed the grid task or polar task, respectively. Narrow bars represent median and interquartile range. Below each stripplot is a bootstrapped distribution of mean differences between conditions in each experiment. (**B**) model LLR per block, separated by condition. Dashed line indicates 0 – the point where either model explains the data equally well. Error bars reflect standard deviations.

4.4.2 Unilateral errors

The results described so far have focused on whether participants generalize correctly or make errors, and how this depends on training curriculum. However, we have not yet characterized what types of errors are made, and whether these are systematic. One option is that subjects learn about both rules in parallel, resembling the type of statistical learning characteristic of artificial learning systems. Alternatively, rule discovery may be independent and relatively abrupt, so that there are periods of time in which subjects have mastered one rule, but not the other. To compare these possibilities, we fitted subject responses using the previously introduced random and bilateral model, in addition to two unilateral models, representing generalization errors that would be made if one, but not both dimensions were learned. We then fitted the sequence of models that best explained each subject's data, allowing models to change between blocks, but only in ascending order of number of dimensions learned, resulting in a maximum of two switch points per subject (full modelling details in the supplementary methods, section b.1.1). The modelling setup is visualized in Figure 4-6a, while Figure 4-6b displays all 191.200 responses made by subjects across our experiments, grouped by their per-block model assignments.

For comparison, we created a simulated counterpart for each subject, whose responses were centered on the ground truth, with a bivariate Gaussian error term, whose covariance was fitted per-block to real-participant generalization accuracies. This error distribution was spherical, i.e. with diagonal covariance and equal variances. This resulted in real and simulated subjects who were matched in per-block accuracy. However, for the simulated subjects, higher error rates by definition corresponded to larger errors in both dimensions, whereas real subjects could in principle learn either in parallel or piecewise. Simulated responses were clipped to the closest edge of the response space if the Gaussian error placed the response out of bounds.

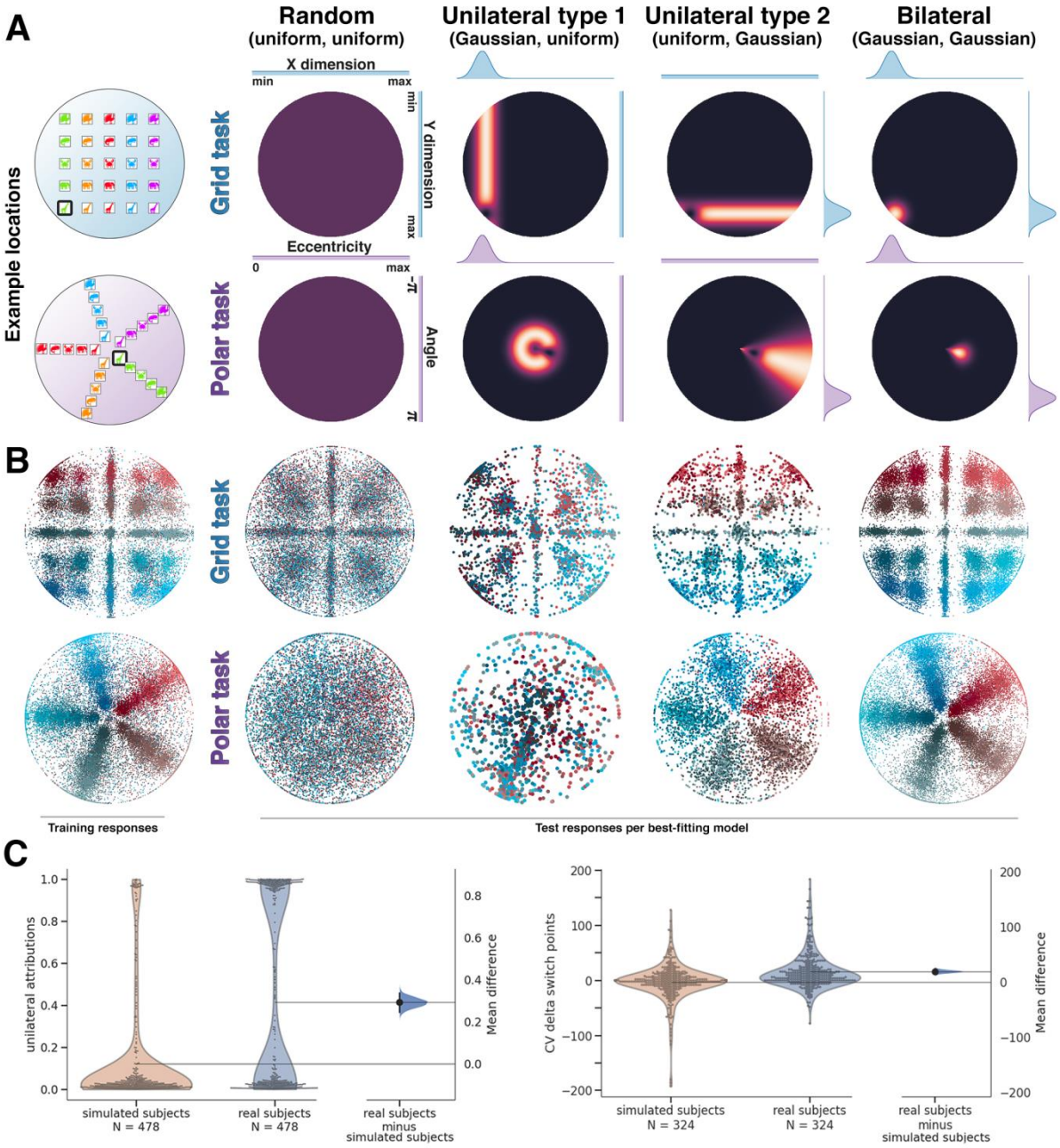

Figure 4-6: Unilateral errors and modelling setup. (**A**) per-model pmfs for one example location (highlighted on the left) in the grid task (top) and polar task (bottom). Dark color represents low probability mass, and bright color represents high probability mass. In our framework, individual rules are either learned (modeled as narrow Gaussian around ground truth) or not (uniform distribution). The resulting model probability mass functions (pmf) are based on the product between per-rule pmfs. With 2 rules per task, this resulted in 4 possibilities. These models were fit to individual subjects, where we allowed switches between models, but only in ascending order of number of dimensions learned. (**B**) left: All feedback location responses. Right: all generalization location responses, grouped according to the model assigned to the corresponding block in our model fitting procedure. Hue indicates y/eccentricity dimension, lightness indicates x/angle

dimension of ground truth. (**C**) left: difference in unilateral model family attributions comparison with simulated subjects, who were matched for accuracy with real subjects, but had a spherical Gaussian error distribution. Each point is a per-subject model family attribution rate. A distribution of bootstrapped differences in means between groups is included, whose black bars represent the 95% confidence interval. Right: crossvalidated difference in per-dimension switch points between simulated and real subjects. Only subjects whose best-fitting model order included the bilateral model were included.

For subsequent testing, we consider that the 7 valid model orders can be grouped into a unilateral model family, consisting of the orders {unilateral, random-unilateral, unilateral-bilateral, random-unilateral-bilateral} and a non-unilateral family consisting of the remaining orders: {random, bilateral, random-bilateral}. Unilateral family attributions were higher in real subjects (median = .044 , SD = 0.455) than in simulated subjects (median = .022, SD = 0.26), $p < .0001$ (bootstrapping test with 10.000 iterations). Convergent with this result, using crossvalidated per-dimension switch point differences (procedure described in the supplementary methods), switch points were farther apart for real subjects (median = 10, SD = 32.5) than for simulated subjects (median = 0, SD = 32.9), bootstrapping test $p < .0001$.

4.5 Discussion

In the current chapter, we set out to study the limits of human transfer, and the factors that promote or impair compositional generalization. Our first contribution is that two factors, axis alignment and blocked training schedules, consistently improved the generalization success of human learners, both in linear and in nonlinear settings. Our second contribution is that the generalization errors humans make are not random, but often contain a 'unilateral' stage, corresponding to consistently correct application of one rule, but not the other. The same was not observed in accuracy-matched control simulations. Comparing human compositional generalization to a vanilla neural network, we confirm our prediction that in the latter, generalization behavior is approximately linear even when the underlying rules are nonlinear. By contrast, human learners attained bilateral generalization in either setting, and in fact generalized better in the nonlinear setting. These findings provide important constraints for learning models and theory, which will be discussed in the next section.

4.5.1 Theoretical implications

With regard to determinants of generalization success, we propose that a common mechanism may explain the benefits of both axis alignment and blocked training. In the category learning setting, learning involves sequential comparisons between current and previous item (Carvalho & Goldstone, 2017), so that blocking highlights within-category similarity, while interleaving highlights between-category differences (Carvalho & Goldstone, 2015). In our task, the fact that e.g. all red stimuli fell on a single line, ring or spoke in any experiment could only be discovered by integrating over multiple trials. If this is discovered primarily through sequential comparisons, it is clear that trial orders in which multiple red stimuli are presented in sequence are preferred. In this view, the blocked, axis-aligned setting is optimal, as successive feedback locations always hold one dimension constant, highlighting the regularity entailed by that dimension. Non axis-aligned training (our 'X' curriculum) provides the greatest departure from this ideal, as it guarantees that no successive feedback locations share a stimulus dimension. Interleaved training forms a more modest disruption, in which shared inputs between successive trials may occur, but are less frequent, and the identity of the constant dimension is less predictable and more variable.

A second result to consider is the presence of unilateral errors. It has previously been suggested that human learners attend a single axis at any given time (Khan, Mutlu, & Zhu, 2011), and that shifts of attention may have theoretical benefits for learning (Kruschke, 2005). In an unstructured representation, such as a vanilla neural network or a parametric error distribution, learning generally proceeds in parallel fashion for different dimensions. Moreover, vanilla neural networks train more efficiently when contexts are interleaved rather than blocked, while the converse applies to humans (Flesch et al., 2018). The finding that humans learn such tasks in a piecewise fashion provides important constraints on the underlying computation. One appealing possibility is that humans use more structured representations, such as one that separates "modules" and a "controller". An exploration of this model type can be found in the section "Cluster-based learning model" in the supplement of this chapter. Such representations are relevant for inference, providing a natural mechanism for generalization, but also have important consequences for learning. Given any prediction error, the system must choose whether to adapt its control system, the content of its modules, or a combination of these. In particular, control errors could give rise to highly structured error

patterns such as unilateral errors, which are not readily accounted for by standard learning models. Our proof-of-concept simulations demonstrate that errors may increase in interleaved and non axis-aligned settings because these are conducive to clusterings that do not match the latent structure. In addition, in the executive control literature, one prominent theory posits that control signals are inert (Allport, Styles, & Hsieh, 1994), explaining phenomena such as switch costs (Duan, Erlich, & Brody, 2015) as interference by residual activation from previous timepoints. Such residual activation may provide an additional mechanism by which control errors are exacerbated when not training in blocked settings.

One limitation of our empirical work is that it is not suited for determining where modules originate. For example, participants learned to generalize the rotation rule in our polar experiments, but it is unlikely that they did so through *de novo* representations of rotation based on the limited number of angles presented in the experiment. Rotation and rotational invariance are ecologically relevant, so we suspect that humans came into the experiment with knowledge of these concepts. Novel tasks can then be learned as flexible recombination of the existing concepts, with unilateral errors representing partial solutions. We expect that functional modules are generally created through integrating information from across many diverse training experiences, a problem we cover in the section "Meta-learning - recursive modular architecture" in the supplement for this chapter.

Another limitation of our work is that we have limited ourselves to settings where the same bidimensional ruleset applied to all trials in an experiment. This is particularly pertinent as the computational framework we suggest involves a controller which is able to flexibly recombine modules. However, as the required rules are the same on every trial, our tasks all belong to a family of problems where the controller need not adapt to different inputs (instead it must solve the module combination problem on the task-level). Our "Cluster-based learning model" assumes that there is a pressure for solutions with low controller demands, but it would be of interest to investigate how these pressures, currently modeled as fixed hyperparameters, may vary as a function of task distribution statistics, following work which suggests that human learners are sensitive to previous task statistics in meta-level generalization (Franklin & Frank, 2020). More generally, we encourage subsequent research to explore settings that place greater demands on control-level flexibility. As an example, a minimal change from our design would be to include shapes which are not animals, signaling a different rule.

Chapter 5: General discussion

Economists claim that we have transitioned to a knowledge-based economy, in which information and learning, not natural resources, form the greatest assets (Houghton & Sheehan, 2000; Powell & Snellman, 2004). Houghton and Sheehan write:

> *"National policies for encouraging knowledge generation, knowledge acquisition, knowledge diffusion, and the exploitation of knowledge have become the most pressing priorities in the science, research and education policy regimes. […] The knowledge economy is seen to demand meta-cognitive skills that are both broad and highly transferable, such as problem-solving and the ability to learn. Knowledge workers are now encouraged to continuously upgrade and broaden their skills, through formal education, lifelong learning, as well as through learning in the workplace and in less formal surroundings. Firms are encouraged to turn themselves into 'learning organisations'"*

Against this backdrop of ever greater demands on our ability to learn, the aim of this thesis has been to shed some light on how learning can be facilitated by using effective training curricula. As previous work has demonstrated rich interactions between the specifics of the learning problem and the nature of the curriculum that promotes learning, it was deemed prudent to focus on the mechanisms underlying curriculum success or failure, to facilitate predictability of the scope within which they remain effective.

The thesis considered three different learning domains and investigated the determinants and mechanisms of curriculum efficacy in each. Whilst the precise curricula were tailored to their respective learning domains, they all operated from the general principle of 'starting small' – restricting the information initially presented to the learner. Other points of focus were how training regimes and the cognitive representations of what is learned interact, and on computational accounts of learning. In this general discussion, I will summarize the findings presented in the thesis, position these in the current literature, and discuss their limitations along with recommendations for future research.

5.1 Overview of results

Chapter 2 focused on the category learning setting, where the manipulation under consideration was distance to the category boundary. The main behavioral result was that human participants learned better when training was restricted to boundary-distal exemplars. Furthermore, this training afforded generalization to other distributions of stimuli, as subjects trained exclusively on boundary-distal examples performed better on boundary-proximal, intermediate and representative distributions of trials than subjects trained directly on these distributions. Computational simulations revealed that this pattern of results can be accounted for by the influence of internal noise, in particular late noise. The interpretation of this mechanism is relatively straightforward: when training closer to the category boundary, a smaller "push" is needed to mistake one category for the other, rendering boundary-proximal learners particularly susceptible to variability in their internal representations. In terms of neural effects, boundary-distal training increased neural sensitivity to the relevant dimension, but did not affect sensitivity to the irrelevant dimension. Furthermore, a neural correlate of prediction error predicted subsequent evidence integration success, suggesting that both objective difficulty (distance-to-bound) and subjective difficulty (prediction errors) influence learning.

Chapter 3 considered probabilistic evidence integration involving multiple cues. The primary manipulation was whether subjects were trained on individual cues or multi-cue combinations, while the eventual test was always on multi-cue combinations. In this setting, no benefit of starting small by training on individual cues was found in terms of differential test accuracies. A minor exception to this was a setting where the total number of cue combinations was low and some were held out during multi-cue training, causing attenuated generalization performance in the multi-cue group. However, this effect did not scale up to a more complex setting with more possible cue combinations. Current theories posit that the central problem in probabilistic evidence integration learning is to learn appropriate cue weightings, but generally do not consider any integration costs when multiple cues must be integrated. As a result, they predict that participants should become more accurate when more cues are presented, as this increases the available evidence. In stark contrast, our results showed that regardless of training curriculum, learners were more accurate on single-cue problems than on multi-cue problems. By adapting the current model to use heuristic decision-making strategies, it was found

that these explain the data better than the standard model in all experiments. In particular, the behavior of participants trained on multi-cue curricula was always best explained by a take-the-best heuristic, which used only a subset of the available cues. Curricula that emphasized the direction of cues over their predictive validity improved the relative fits of the tallying model, in which decisions are based on the sum of cue signs.

Finally, Chapter 4 considered a compositional rule learning task, in which two separate rules had to be inferred from a limited number of input combinations. Here it was found that participants benefitted from training on one rule at a time, holding the other constant. Furthermore, it was observed that the error patterns they exhibited were highly regular and 'unilateral', containing errors in the application of one rule, but not the other. These results were compared with feedforward neural network models, which have previously been found to benefit from interleaved training. Moreover, we showed that these networks only succeeded at generalization when the underlying rules were linear, while human learners reproduced the ground truth generalization pattern regardless of whether rules were linear or nonlinear. To explain these findings, we further developed the theory that human thought is inherently compositional, involving clustered, recombinable representations. Simulations embodying these principles captured several of the phenomena exhibited by human, including a benefit from blocked training, the presence of unilateral errors, and an ability to generalize nonlinearly.

5.2 Relation to existing work

The work discussed in this thesis is not presented against a blank slate. Some training schedules and the mechanisms subserving their success are already well-characterized. Best known and studied among these is that for problems which permit division into successive approximations, shaping procedures accelerate learning at least in part by mitigating the curse of dimensionality (Sutton & Barto, 2018). Our focus is on the large subset of learning problems for which the current literature does not provide ready answers on how they can best be learned.

In attempting to answer this question, several of our contributions have borrowed well-developed theories from other domains, and considered their ramifications on learning. For example, a broad body of work describes that human representations are inherently noisy, so that their internal estimates

vary from trial to trial, even when inputs are identical. Our contribution is that this has implications for learning of parametric categories, as internal noise disproportionately affects learning when boundary-proximal examples are used.

It has also been widely observed that human behavior is well-characterized by heuristic decision-making strategies across a wide range of settings. Interestingly, what consequences this may have for learning is a topic that has received minimal attention, while the core tenet of these heuristics – that they discard part of the available information – naturally raises the question how information can be learned if it is thrown away. For example, under the take-the-best heuristic (Gigerenzer & Goldstein, 1996), only the most valid cue is used, while all other cues are discarded. Clearly, this requires knowledge of the relative validity of each cue. But if lower-validity cues are systematically discarded, how can their validity be learned? In this regard, one of our contributions is that human behavior is better explained when assuming that this heuristic affects learning, and only the cues used for decision-making are updated. This suggests an interaction between learning and subsequent information sampling, even in full-feedback settings.

We have also considered the benefits of blocked training in the context of a compositional rule learning task. Whilst our result was obtained in this specific setting, an increasingly lucid, domain-general picture is emerging which suggests that when learning requires integration across different examples within contexts (Flesch et al., 2018), categories (Carvalho & Goldstone, 2014), dissociable rules, etc., human learners benefit from temporal blocking of the elements within which integration must take place. One possibility is that time provides an intrinsic contextual signal, guiding learning. Supporting this view, when participants are trained on two different sets of items on different days and then are then queried on both sets during a third session in an fMRI scanner, the day on which each item was learned can accurately be decoded from the BOLD signal, and in a post-task questionnaire, participants are able to accurately report on which day they learned each item, even if this has no bearing on the task (Balaguer, 2018).

Our final contribution was that several aspects of human compositional learning can be accounted for by a system that separates what it learns into distinct clusters, along with a control mechanism that

combines these. There is a long history describing that connectionist architectures do not readily account for compositional structure, spurred on by an early critique that classical logic, but not connectionist theory, permits a 'combinatorial semantics' (J. A. Fodor & Pylyshyn, 1988). This is contrasted with symbolic approaches, which use abstract compositional representations whose elements represent objects and relations. Notably, the strengths of symbolic approaches form a close match with the limitations of connectionist approaches (Garnelo, Arulkumaran, & Shanahan, 2016) , and attempts are ongoing to imbue connectionist approaches with symbolic features, aiming to get the best of both worlds (Garnelo et al., 2016; G. F. Marcus, 2003; Nye et al., 2020). To some extent, our models (sections b.3 and b.4) may be viewed as belonging to this class of hybrid models, where controller outputs provide a symbolic description of the solution. The interpretability of this solution will depend on the interpretability of the clusters themselves (e.g. “rotate by X, then scale by Y”). Fortunately, as generalization success depends on the degree to which clusters match the generative structure of a problem, this aligns the goals of performance and interpretability.

5.3 Limitations and prospects for extension

A central motivation for the work in this thesis and the focus on mechanistic explanations was the observation that in past work, the same curricula were effective in some settings, but not in others, making it difficult to predict when a given curriculum would succeed. With this in mind, it seems relevant to consider the scope under which our findings are likely to hold. The finding that curriculum efficacy relates to internal noise is not limited to category learning per se, but a likely requirement is that input dimensions are continuous. Moreover, for differential curriculum effects, the relation between inputs and outputs must contain some form of nonlinearity whereby in some regions of the input space, small changes in inputs incur large changes in outputs, such as a boundary between categories or an isopreference curve between decision-making alternatives. This structure enables the identification of regions that are sensitive to internal noise. In addition, determinants of noise magnitude must be considered. For example, in one-dimensional category learning tasks using simple shapes of different lengths, no benefit of boundary-distal training is observed (Pashler & Mozer, 2013). A relatively unexplored question is to what extent attention affects internal noise. Potentially, techniques for guiding

attention toward relevant features could form an alternative for mitigating the deleterious effects of noise.

Given the large number of conceptual replications performed, the finding that training on single (probabilistic) cues does not benefit learning compared to training on multi-cue examples given sufficient training is firmly established. However, our design provided limited insight into whether this manipulation affects the rate of learning, and it remains possible that on shorter timescales, the number of simultaneous training cues matters. While human responses were consistently better explained by heuristic decision-making strategies than by normatively optimal integration, we have considered only a limited candidate model space, and other possibilities, such as limited cue capacity without ordering by validity, for example using an attention mechanism that mediates which cues enter consideration, remain well within the realm of possibility. Furthermore, we returned from our short foray into the consequences of heuristics on learning with testable predictions for future research. For example, if a take-the-best strategy constrains learning, this suggests that cue learning should be subject to context effects, whereby how well an individual cue is learned should depend on the validity of the cues that are presented alongside it. Another point that may warrant further attention is that the nature of the training curriculum predicted which heuristic was used in the subsequent test phase, even though this phase was identical in both conditions, and full feedback was provided for 400 trials. This tentatively suggests that choice of decision strategy may be particularly sensitive to the conditions during initial training. The precise dynamics of decision strategy change over time could be clarified in further research.

In modelling human compositional generalization, we have relied upon systems with modular knowledge. One open problem is where these modules come from in real human learners. While we demonstrate that computationally, general functions (such as rotation) can be learned from a limited number of examples (such as a low number of orientations) in many different settings, this is expected to be a lengthy process, making it difficult to verify in humans. A second limitation is that our focus has been just on a single type of compositionality (“systematicity”, Hupkes, Dankers, Mul, & Bruni, 2020) and does not address the other forms. In all compositional learning tasks we studied, stimulus dimensions (e.g. color) determined a rule type, while dimension levels (e.g. red) indicating a rule

magnitude. This gives the task a hierarchical structure. For example, if a new color were introduced, the appropriate output could not be known immediately, but the possible solution space is already highly constrained by its superordinate category. While some research has focused on how such hierarchical structure is learned by humans and how it can be utilized in computational models (Balaguer, 2018), more research is required to fully understand this process.

5.4 Conclusion

The main aim of this thesis was to contribute to our understanding of the determinants of learning success. My doctoral work has assessed the premise that restricting initial training to an appropriate set of examples promotes learning in a variety of settings. This was tested through rigorous experimentation and computer simulations, involving both off-the-shelf and adapted models. Taken together, I hope my work provides a small step towards understanding how we can learn more effectively, and potentially spark inspiration in endeavors such as bringing the successes of human learning to machines, and improving our educational regimes.

Supplementary information for Chapter 2

b.1 Supplementary methods

b.1.1 Abstract task representation simulations

The network was trained and tested on flattened 50x50 inputs. The central 26x26 of these contained the parameter values used in the experiment, while the remainder was used to minimize clipping due to covariance or late noise. Activations were proportional to the probability of a bivariate Gaussian distribution, whose mean was determined by the per-trial stimulus parameters, and whose covariance was fitted per subject. For the network architecture, we used a 1-hidden layer MLP with 64 hidden units, and a single output unit reflecting choice probability. ReLu nonlinearities were used in the hidden layer, and a sigmoid function in the output layer. We used He initialization (He et al., 2015) and optimized with stochastic gradient descent, computed on the cross-entropy between model outputs and ground truth.

b.1.2 Model fitting procedure

To establish how the simulations relate to human behavior, the abstract task model was fitted to individual subjects' responses. Best-fitting parameters were obtained per subject by a grid search procedure over learning rate, relevant variance and irrelevant variance in the noiseless setting. Then, search was performed over learning rate and noise magnitude, under the variances obtained in the noiseless setting. This was done separately for each noise type, in order to assess which best accounts for human behavior. A simulation was run 10 times for each combination of these factors for each subject. In each run, the weights of the model were randomly initialized, and noise was resampled (note that the noise magnitude parameter determines the scale of the noise, but the effective per-trial amounts vary from run to run, as noise is randomly sampled per trial). Mean variance in per-trial response probability was 0.036 between runs. Then, for each subject, the parameters were

selected which minimized the squared error between mean simulated response probabilities and empirical responses.

Subsequently, we ran a fixed-effects Bayesian model selection analysis (Stephan, Penny, Daunizeau, & Moran, 2009) as implemented in the VBA toolbox (Daunizeau, Adam, & Rigoux, 2014) to calculate exceedance probabilities. To chart the learning dynamics more generally for different noise types, noise magnitudes and curricula, the model was trained using a fixed covariance and learning rate (the medians obtained in our model fitting, pooling over groups). Models were trained for 500 cycles on 20 levels of noise magnitude per noise type. On each cycle, the trial structure was randomly generated (following the same restrictions as in the main experiment), random starting weights were used and noise was resampled.

b.1.3 EEG acquisition

The EEG was recorded using 61 Ag/AgCl sintered electrodes laid out according to the extended international 10-20 system. Data was recorded at 1000Hz using a Compumedics Neuroscan SynampsRT amplifier. Electro-oculography was recorded using electrodes lateral to either eye, and above and below the right eye. Data was referenced online with respect to the right mastoid. Data was downsampled to 250Hz and put through a 1-40Hz bandpass filter for preprocessing. Average referencing was then performed on all electrodes. Faulty channels were selected based on visual inspection and interpolated from surrounding activity. If any channels were interpolated, previous pre-processing steps were repeated. The data was then epoched, after which trials with visible artifacts were removed. Finally, further artifacts were removed using individual component analysis. In all analyses, we subtract a baseline signal from each electrode, which is its average activation in the interval from 500 to 0ms before stimulus onset, or 500 to 0ms before feedback onset for analyses pertaining to feedback signals.

For our analyses involving CPP, we use a centro-parietal ROI defined as the electrodes CPz, CP1, CP2, Pz, P1 and P2. Our CPP amplitudes are defined as the average voltage at these electrodes over the time window of 308 to 792 ms after stimulus onset. These sites and this time-window were determined using a cluster-based method (Maris & Oostenveld, 2007) on all subjects and taking the

cluster with the highest T-statistic between 0 and 1000ms, using a pointwise inclusion threshold of p=.001. For commensurability between conditions, only test trials were included. These sites overlap with those reported in previous work (Herding, Ludwig, von Lautz, Spitzer, & Blankenburg, 2019; Luyckx, Nili, Spitzer, & Summerfield, 2018). As prescribed in Kelly & O'Connell (2014), we performed a Current Source Density transformation on the neural data to reduce overlap with frontocentral motor negativity, which otherwise corrupts CPP amplitude measurements, especially when it has a late offset.

As the FRN manifests as a negative inflection that overlaps spatially and temporally with the feedback-induced P300, base-to-peak measurements are deemed more reliable than time-window averages (Hajcak, Moser, Holroyd, & Simons, 2005; Holroyd, Nieuwenhuis, Yeung, & Cohen, 2003; Pfabigan et al., 2011). We used the signal at Fz, following Huang & Yu (2014) and computed the maximum between 150 and 350 ms, the minimum between this maximum and 400 ms and the maximum between said minimum and 450 ms. FRN magnitude was defined as the average of both base-to-peak measures.

b.1.4 Determinants of the CPP signal

The CPP signal was expected to reflect decision certainty. However, as our design interleaves training and test trials, behavioral and neural tuning to dimensions was unlikely to remain static over the course of the experiment. To capture this flexibility, we estimated psychometric tuning at every test trial based on the surrounding 120 trials, leaving out that trial. A 4-parameter sigmoid function was fit to the data in this time window (Eq. a.1). The window size was determined by performing exhaustive search on window sizes (in steps of 10 trials) and selecting the window that minimized the squared error between model choice prediction and subject response.

$$\hat{p}_{right} = f(x|a,b,c,d) = a + \frac{b}{1 + e^{\frac{-(x-c)}{d}}} \quad \text{(Eq. a.1)}$$

This process was performed once for the relevant dimension and once for the irrelevant dimension. The *x* term in this formula represents the dimension value (i.e. the ground truth degree of branchiness

or leafiness), which was bounded between 0 and 1. Eq. a.1 is also used in section 3.8, where *x* represents the integrated choice probability *p(choice|right)*.

In chapter 2, the signed tuning curve $f(x)$ obtained in this way was transformed into an unsigned tuning curve $g(x)$ (dropping the dependency on a, b, c and d for legibility):

$$g(x) = \mathrm{abs}(f(x) - f(c)) \qquad \text{(Eq. a.2)}$$

To obtain a learning covariate, a powerlaw model was fit to the per-trial accuracy data for each subject. These three factors and first-order interactions with learning were then z-scored and used as inputs for the regression model. A full overview of the resulting weights, split up by condition, is visualized below in Figure a-1.

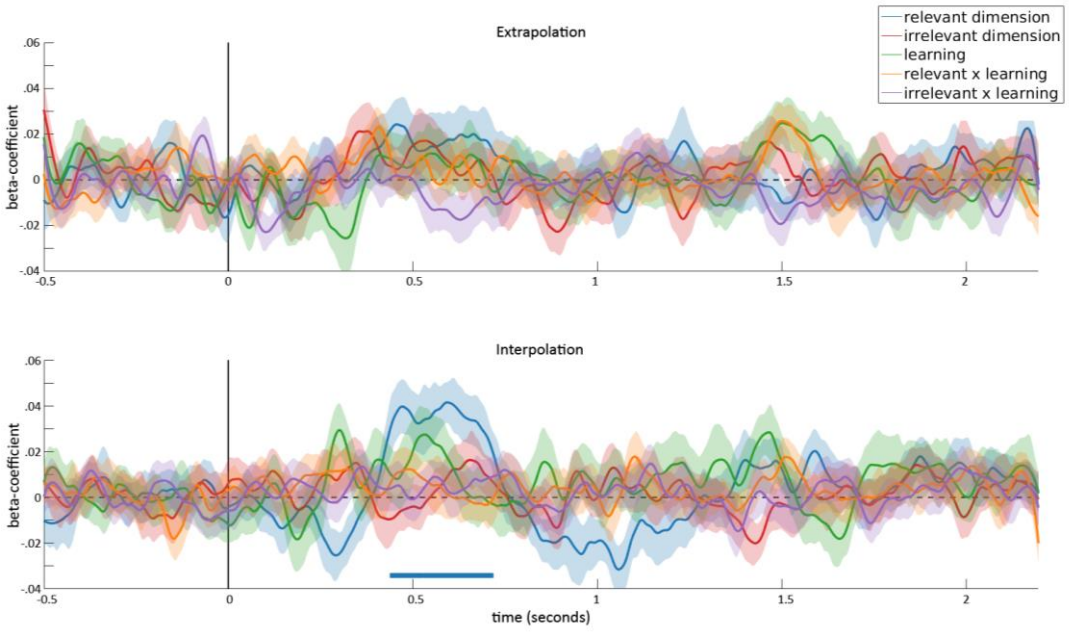

Figure b-1: Per-timepoint psychometric regression analysis, split up by condition. Plotted is time in milliseconds against regressor beta coefficients. The blue line indicates local relevant dimension tuning, while the red line indicates local irrelevant dimension tuning. The green line represents a power-law function fitted to per-subject accuracy. The orange and purple lines represent interactions between learning and the relevant or irrelevant dimension, respectively. Solid bars at the bottom of the graph indicate pointwise t-test significance at p<.05. Stimulus onset was at 0ms, and offset was at 1000ms.

b.2 Additional online experiments

Experiment 2 was an online replication of the main experiment, with the addition of the intermediate condition detailed in Figure 2-3. All details regarding trial structure and stimuli are exactly as in the EEG experiment from the main text. Experiment 3 is a conceptual replication and extension, in which we allowed curricula to change over time and compared different irrelevant dimension curricula. After every 2 blocks, the distribution from which parameters were drawn changed, either decreasing (far-near, FN) or increasing (near-far, NF) in mean distance-to-bound. This curriculum design is visualized in Figure a-2a. Curricula were assigned independently for the relevant and irrelevant dimension, resulting in a 2x2 factorial design, in addition to a uniform control condition. To implement the curricula, parameters were drawn from a concatenated beta distribution, which was defined as follows:

$$pdf(x) = \begin{cases} 2\mathrm{x}^{\alpha-1}(1-2\mathrm{x})^{\beta-1}, 0 \leq \mathrm{x} \leq 0.5 \\ 2(\mathrm{x}-0.5)^{\beta-1}(2-2\mathrm{x})^{\alpha-1}, 0.5 < \mathrm{x} \leq 1 \end{cases} \quad \text{(Eq. a.3)}$$

Using this design, we observe no significant difference in slopes between FN and NF in the irrelevant dimension (U=2152, p=.087) and so pool over irrelevant dimensions for subsequent analyses and in the main text. In the relevant dimension, FN slopes were higher than uniform slopes, which were in turn higher than NF slopes (p<.0004 for all paired comparisons). Fitting only to the hardest 5-35% of trials, FN curriculum slopes remained higher than NF curriculum slopes (U=1363, p<.0001). These results are visualized in Figure a-2b.

Crucially, in both additional experiments, the interpolation group learns faster than the respective control groups, even when evaluated on the control group's training distribution. This rules out the possibility that the interpolation benefit hinges on the boundary-proximal condition being sufficiently difficult that the task is never fully learned, as both the intermediate and uniform conditions reach comparable asymptotic accuracy levels to the interpolation condition.

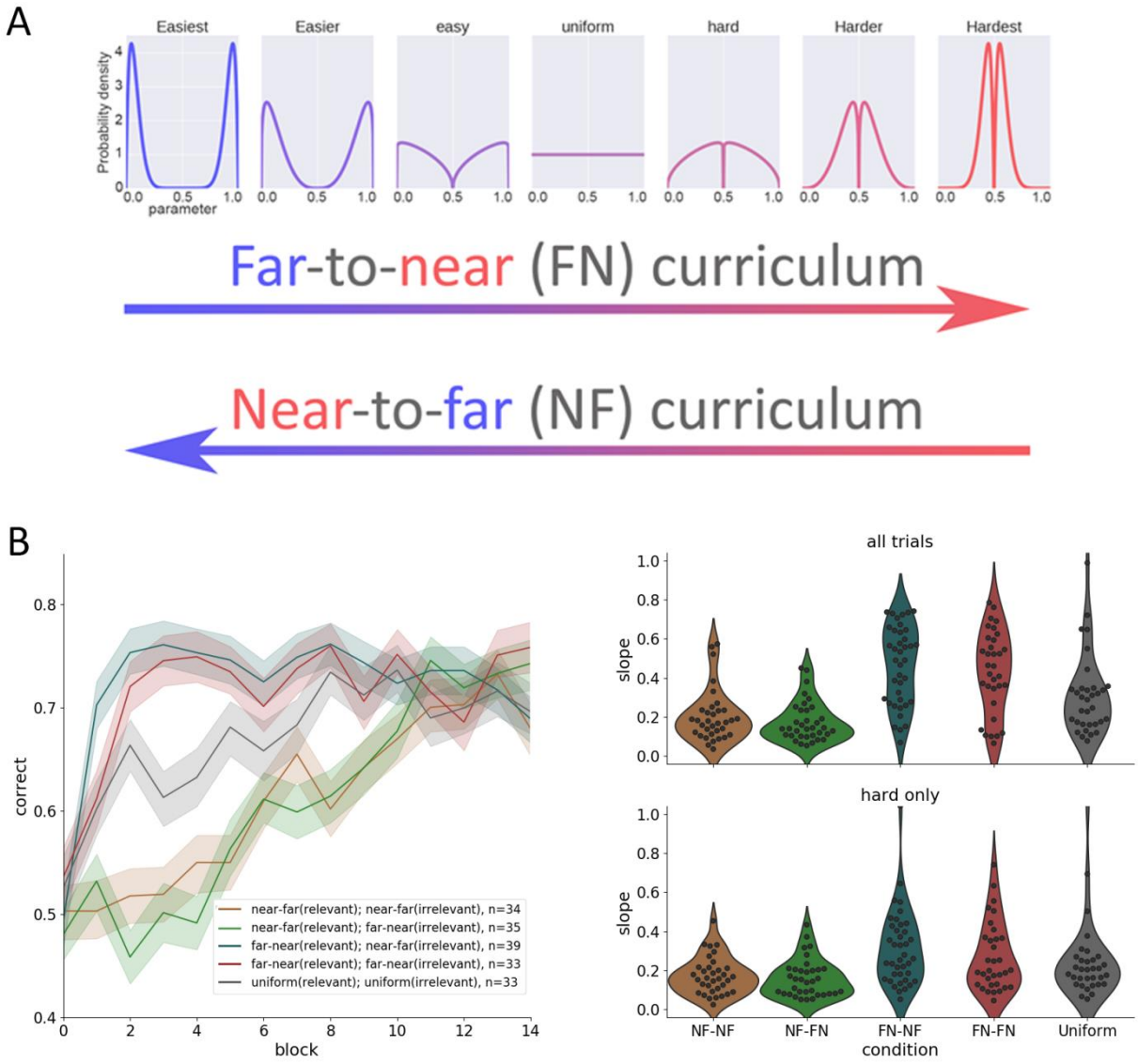

Figure b-2: Experiment 3: detailed setup and results. (**A**) Training trial distributions, which changed every 2 blocks, either starting far from the boundary and progressively sampling closer (far-near, FN) or vice versa (near-far, NF),. This curriculum was applied factorially to the relevant and irrelevant dimensions, yielding 4 combinations, in addition to a full uniform control condition. (**B**) Behavioral results. (**Left**) Learning curves, displaying per-block mean test trial accuracy, split up by curriculum. (**Right**) Violinplots, displaying per-subject power-law slope, fitted to test trials in the first half of the experiment. Plotted are fits to all test trials, and below that fits to only the hardest 5-35% of test trials.

b.3 Model fitting

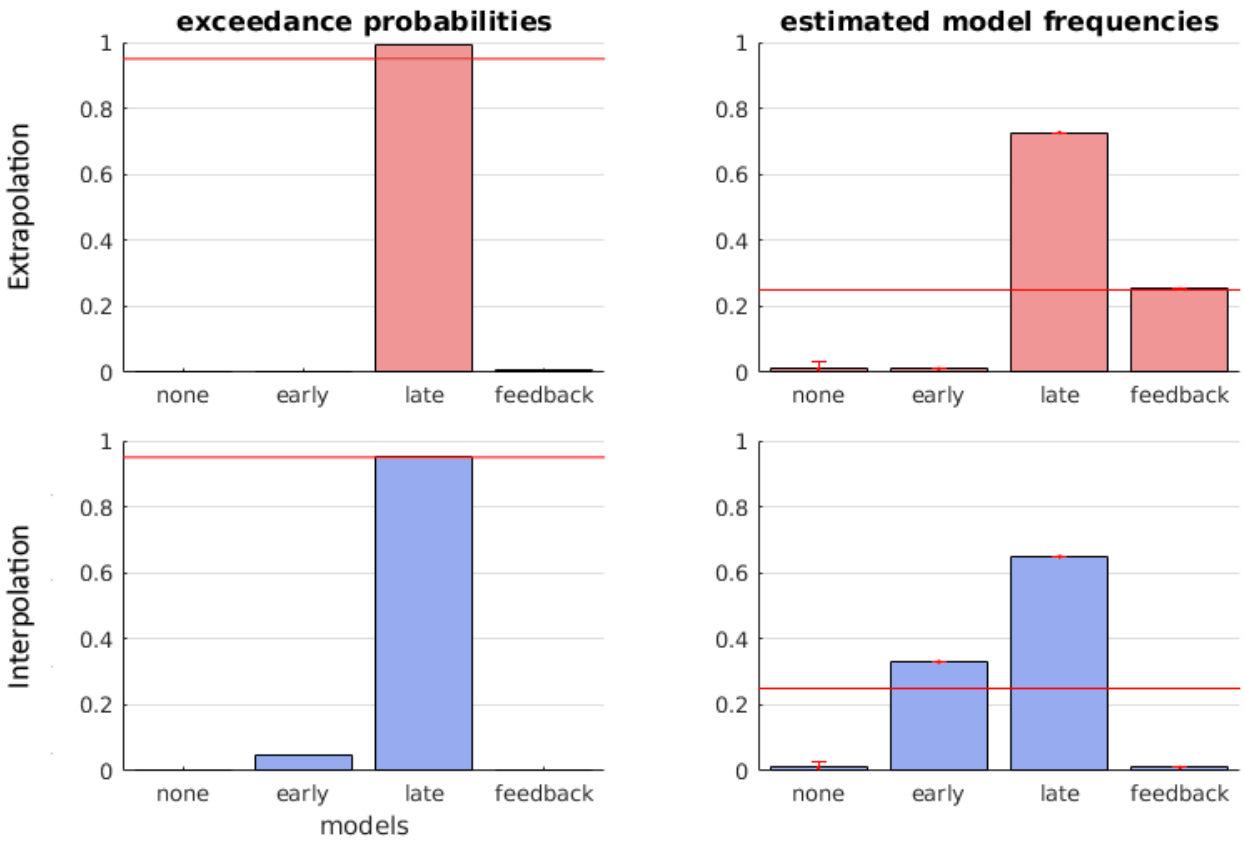

Figure b-3: Per-condition outcomes of Bayesian model selection procedure. Compared are the no noise, early noise, late noise and feedback noise models from the main text. Results are separated by condition (Top: extrapolation, Bottom: interpolation). (**Left**) Per-model exceedance probability. Red line indicates 95% significance threshold. (**Right**) Model attributions. Red line indicates 25% chance level.

b.4 Crossvalidating orientation for psychometric analyses

In this section, we will describe our method of fitting psychometric functions as well as demonstrate with simulated data why this is important in a design with one or more irrelevant dimensions and multiple subjects. The core concept behind this approach is that the directionality of sensitivity to a relevant dimension (that is, whether a certain response becomes more or less likely as that dimension increases) is generally systematic between subjects, as it is tied to trial outcome. However, how choice depends on the irrelevant dimension may instead be asystematic between subjects, depending on idiosyncrasies in trial order and participant priors. In that case, they may be lost in the average, even when sensitivity is steep within individual subjects.

The fitting procedure was as follows. For each subject, test trial data were divided into 2 folds, one for even and one for odd trials. We then fitted Eq. a.1 to each subject within folds, with a positivity constraint on *b*. If *a* was greater than *d* in either fold, they were swapped in the other. This creates two sets of parameters per subject per dimension. For plotting purposes, the average of plots generated by each set of parameters was then computed per subject, matching the original number of units of observation. Conceptually, this manipulation places psychometric sensitivity in the frame of reference of the response that becomes more likely when a dimension is increased, where the identity of that response category is estimated in crossvalidation to prevent circularity or inflated estimates.

To create simulated data for demonstration purposes here, we again used Eq. a.1. For each simulated subject, parameters were drawn for a separate relevant and irrelevant tuning curve. Parameters were sampled uniformly, with the following constraints. We constrained the maximum parameter (*d*) to be above the minimum (*a*) for the relevant model, but did not impose this constraint for the irrelevant model. We constrained the slope (*b*) to be positive. In the conditions where the irrelevant dimension was not used, this was implemented by setting both *a* and *d* to 0.5. In total, we simulated 1000 subjects, each having 280 trials, the number of test trials in the second half of the real experiment. For each trial, choice probabilities from the relevant and irrelevant tuning curves were averaged, and responses sampled from this integrated response probability. Results from fitting to this simulated data are displayed in Figure a-4. Indeed, we observe that using the standard method, similar results are obtained when the irrelevant dimension is not used as when it is used as much as the relevant dimension, but with an inconsistent direction between subjects. When the irrelevant dimension is not used, our cross-validated directions approach correctly retrieves the same results as the standard method. However, with cross-validated directions, we additionally become able to retrieve sensitivity to the irrelevant dimension when its orientation is variable between, but consistent within subjects.

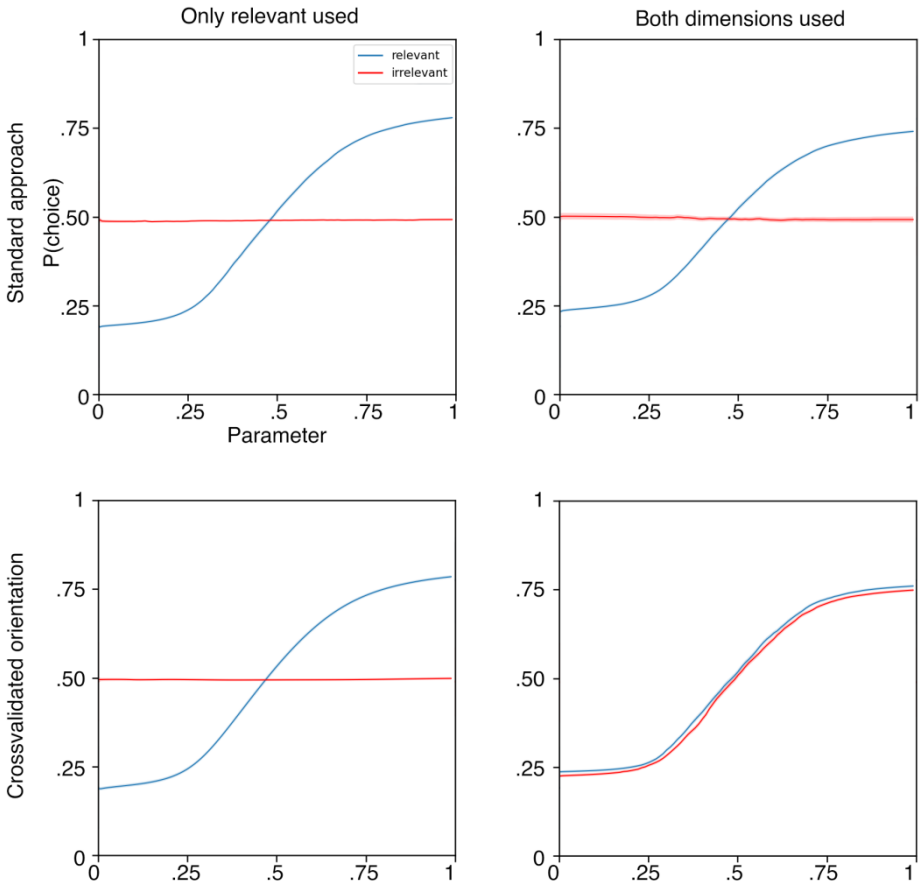

Figure b-4: Crossvalidated orientation psychometric model validation on simulated data. All plots visualize the effect of parameter magnitude on choice probability. The relevant dimension is displayed in blue, while the irrelevant dimension is in red. (Top Left) when only the relevant dimension guides response, the standard model accurately captures this. (Top Right) when the irrelevant dimension has as much bearing on response as the relevant dimension, but its directionality differs between subjects, the standard approach does not pick up on this. (Bottom Left) When crossvalidating orientation and only the relevant dimension is used, this approach yields similar results to the standard method. (Bottom Right) When both dimensions are used, the method that estimates orientation in crossvalidation is able to accurately recover sensitivity to the irrelevant dimension.

b.5 Convolutional neural network simulations

In the main text for this chapter, we investigated curriculum effects using abstracted inputs and a simple model. Whilst this allows for ease-of-manipulation and greater interpretability, an open question is how different learning curricula affect the dynamics of full-scale classification models. In addition, the task subjects face involves a perceptual component, in which they must infer the relevant dimension quantity from complex visual inputs. This perceptual component is abstracted away by modelling the inputs as bivariate Gaussians.

b.5.1 Design

In the current section, we describe an extension where learning was simulated using the same stimuli that real subjects faced, using convolutional neural networks. As these networks are adept at processing two-dimensional stimuli such as images, we trained them on the same stimuli as real human subjects. Unlike humans, however, these networks train solely via a straightforward optimization procedure free from human priors and biases. Thus, a correspondence in observed curriculum effects would be informative of a purely computational mechanism subserving curriculum effects in the absence of noise and prior beliefs. To ensure our findings hold in full-scale models, we used a state-of-the-art convolutional neural network architecture which had been pre-trained on millions of naturalistic images (AlexNet; Krizhevsky, Sutskever, & Hinton, 2012). For comparison, we also trained a simpler convolutional neural network from randomly initialized weights. In the pre-trained network, weights in the initial two layers are frozen and learning takes place exclusively in the more abstract later layers, allowing us to gauge whether the element of perceptual learning is instrumental in obtaining curriculum differences.

b.5.2 Procedure

An overview of the randomly initialized CNN architecture is given in Figure a-5. For the pre-trained architecture, we used AlexNet (Krizhevsky et al., 2012).

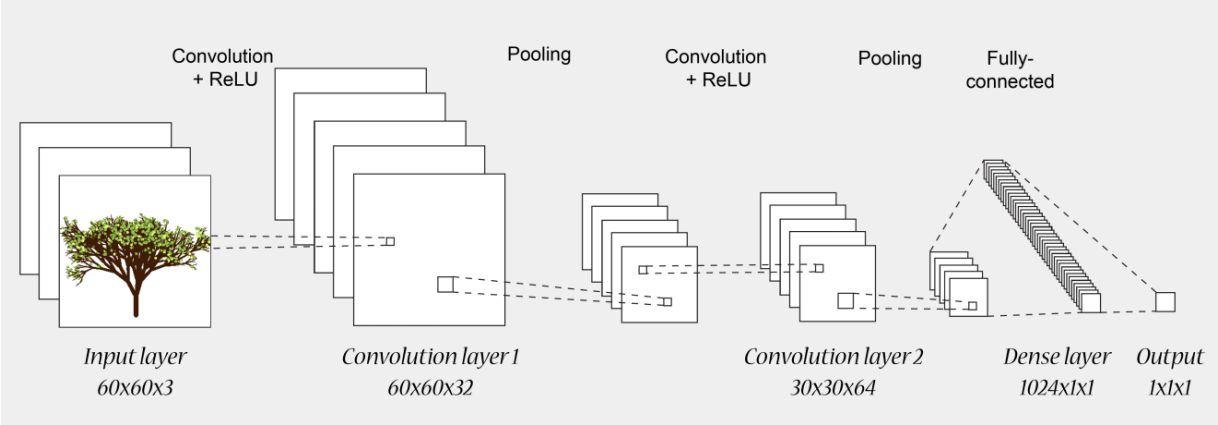

Figure b-5: Randomly initialized convolutional neural network architecture. Inputs were 60x60 RGB images. These were processed by two convolutional layers, each using a rectified linear unit (ReLU) activation function and 2x2 max pooling. Then, the signal passed through a 1024-node dense layer, before producing a single output.

The CNNs were trained and tested on 60x60 RGB images. We ran 100 cycles per condition. In each cycle, the network was trained on a single epoch, consisting of 8000 training examples in minibatches of 100 trials. After each minibatch, the network was evaluated on the full test set of 1001 examples, during which weights were not updated. After completing all trial in a cycle, the order of training examples was randomized, weights and biases were reset, and the next cycle began. In these simulations, the relevant dimension was always the branchiness of the trees. A sigmoid function was used in the output layer. We used He initialization (He et al., 2015) and optimized with stochastic gradient descent, computed on the cross-entropy between model outputs and ground truth.

b.5.3 Results

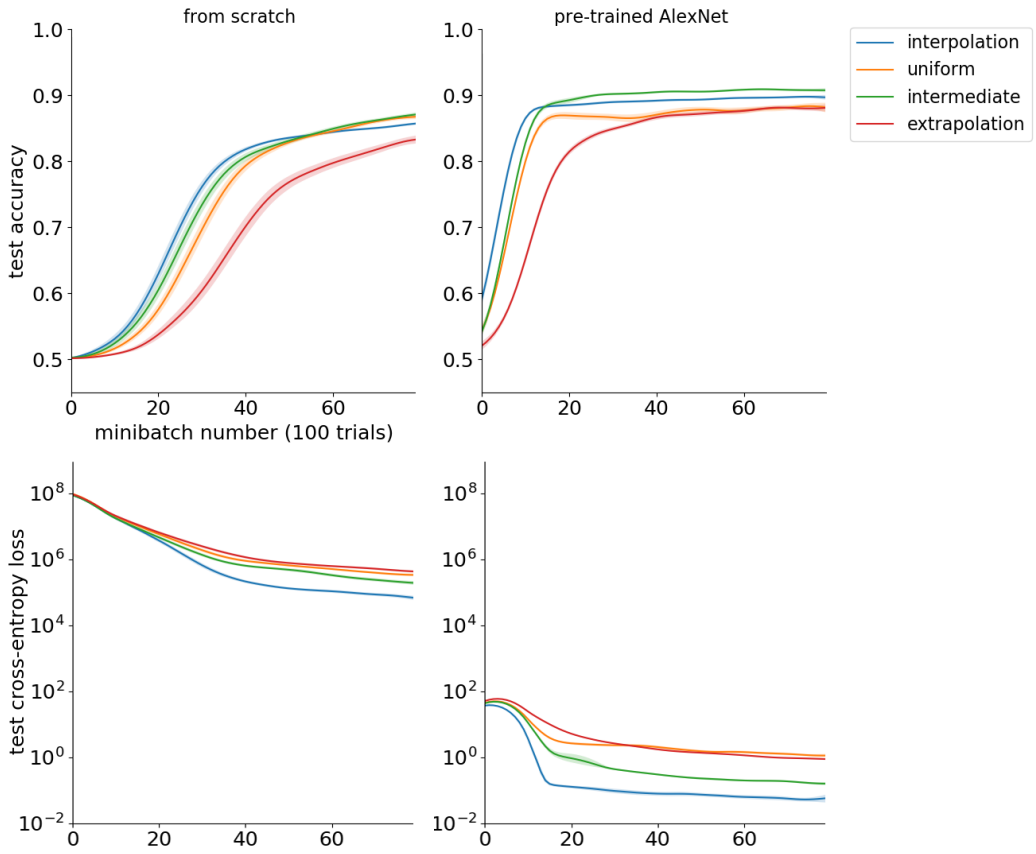

Figure b-6: Effect of curriculum on learning in deep convolutional networks. Learning curves show performance on uniformly drawn test exemplars, for which no gradient is computed. Accuracy is plotted against number of training minibatches. Left: randomly initialized network. Right: pre-trained network (AlexNet) with layers 1-2 frozen. Error bars represent 95% confidence intervals (CIs) over 100 cycles.

Learning curves for both networks are plotted in Figure a-6. As in humans, both the randomly initialized model and the pre-trained AlexNet model exhibited a benefit from training on more diagnostic exemplars. These findings are in concordance with other empirical work (Hacohen & Weinshall, 2019) and suggest that a more fundamental dynamics contributes to curriculum effects, over and above the effects related to internal noise in humans and other organisms.

Supplementary information for Chapter 4

c.1 Supplementary methods

c.1.1 Model fitting

As responses were measured at specific pixels, our probabilistic modelling took place in discrete space. The valid response space was a circle with a radius of 265 pixels. This was implemented by first defining probability distributions in a 530x530 square response space, then masking out locations falling outside the circle, then normalizing probability mass by the remaining sum to account for clipping effects and discretization. In the grid tasks, we used a 2D meshgrid representing X and Y locations of pixels, centered on the screen center. These values were then divided by 180, so that ground truth locations fell on the values {-1,-0.5,0,0.5,1} of their corresponding dimension, and the outer edge of the response space corresponded to a Euclidian norm of 265/180 ≈ 1.472. For the polar task, X and Y values in the mesh grid were converted into polar coordinates (angle and eccentricity). The random model was obtained by assigning the same probability to each valid location, which was simply the reciprocal of the total number of locations. The bilateral model was obtained by evaluating a Gaussian distribution separately on the X and Y meshgrids (angle and eccentricity for polar), masking, normalizing, taking the product of both distributions, and normalizing again. To unpack this with an example, in the grid task, a Gaussian for the X-dimension would be computed by evaluating a Gaussian function on the X meshgrid. This was then masked (discarding locations outside of the circle) and normalized (to make the probability distribution sum to one). We call this a 'raw unilateral model'. Then, this was multiplied element-wise with a similar distribution evaluated on the Y-meshgrid and normalized again, resulting in the bilateral model probability mass function. To ensure unilateral distributions modeled errors rather than correct generalization with greater tolerance, raw unilateral models were multiplied by a validity map which was inversely proportional to the bilateral distribution. Means of Gaussians were the ground truth coordinates. Standard deviations corresponded to 1/20th of

the total space. For x, y and eccentricity, this was .147. For angle, this was $\pi/10$. To allow for occasional deviations from systematicity, a lapse rate of 1% was built into each model by using a weighted average of 99% of the main model and 1% of the random model for each model.

As the identity of the model that best explained the data was not expected to remain the same over the course of learning, we allowed model choice to change over time, but only in ascending direction of dimensions learned, resulting in a maximum of two switches per subject. Only generalization trials were fitted. In determining model order, switches were constrained to only take place between blocks. The primary reason for this is that a priori, we deemed between-block switch points most plausible, as these encompass all feedback trials. In addition, this constrained assigned models to remain stable for at least 16 trials (the number of generalization trials per block), reducing the risk of overfitting. BIC was used to penalize the additional complexity entailed by multi-model orders: $BIC = -2log\big(p(data|model\ order)\big) + N_{switches} \cdot log(16 \cdot 14)$. Subsequently, we used Variational Bayes for model selection (Stephan et al., 2009) as implemented in the VBA toolbox (Daunizeau et al., 2014) to calculate per-subject model order attributions. The inputs for this analysis were $LL' = -.5 \cdot BIC$, the complexity-adjusted log-likelihoods of each model order for each subject.

c.1.2 Cross-validated delta switch point analysis

The goal of this analysis was to estimate per-dimension switch points, so that one can then determine how many trials apart either dimension was learned. However, with naïve application of this procedure, as switch points cannot be determined with absolute precision, some difference in switch points between dimensions would be expected even if there were no effect. For this reason, switch points were determined in 2-fold crossvalidation, using one fold to determine the expected direction of the effect and the other to quantify the difference in switch points in that frame of reference. This resulted in positive values for consistent differences between folds, and negative values for inconsistent directions between folds. Switch points were estimated per dimension per fold with a 4-parameter sigmoid model (Eq. a.1), taking the inflection point parameter to be the switch point. This resulted in two estimates per subject, the difference in switch points between dimensions in fold 1 normalized by the direction of fold 2, and the difference between dimensions in fold 2 normalized by the direction of

fold 1. These were condensed into a single per-subject average by taking their average. Only generalization trials were used in this analysis, and the two folds were defined as the odd versus even remaining trials.

c.1.3 Vector addition model

In the introduction of this chapter, we motivated the prediction that a canonical neural network would generalize according to a linear solution when possible. To test this empirically, we first analytically computed a linear solution to our tasks. Specifically, a linear function must obey the identity $f(x + y) = f(x) + f(y)$. Using 1-hot inputs per dimension, this implies that the output corresponding to each input dimension is a vector, and the composition of both dimensions is simply the sum of these vectors. In Figure 4-4, a solution satisfying this constraint is given by first determining the centroids of one dimension, and then the vector that must be added to these centroids to account for the second dimension.

c.1.4 Canonical neural network simulations

For the canonical neural network simulations displayed in Figure 4-4, inputs were length-10 2-hot vectors representing shape and color. The network contained a single hidden layer, comprising 64 nodes. A sigmoid nonlinearity was used in the hidden layer. The output layer was linear, projecting onto 2 nodes, representing the model's x and y location predictions. The output scale was as in the model fitting procedure. The learning rate was .05 and bias nodes were used in the hidden and output layer. The Adam optimization algorithm was used, computed on mean squared error with an l2-penalty of 0.0001 as the loss function. The batch size was 1 trial. Locations visualized in Figure 4-4 are mean predictions for all inputs computed at the end of 10.000 runs per task, resetting the weights using Xavier initialization (Glorot & Bengio, 2010) and generating a new valid training order at the beginning of every run.

c.2 Effect of subject priors on learning

Recording a stimulus dissimilarity rating task (for brevity, "arena task" in the remainder of this section) before the experiment allowed us to gauge subjects' prior beliefs about the stimulus structure. As no

feedback was provided and the used colors and shapes differed from the main task, it was not possible to apply the model fitting setup, or to match exemplars from the arena task with those from the main task. Instead, arena responses were classified using representational similarity analysis (Nili, Wingfield, Walther, & Su, 2014). For each subject, an empirical representational dissimilarity matrix (RDM) was defined as the Euclidian distance between each pair of stimuli. Model RDMs were constructed for random, unilateral and bilateral models. In the random model, dissimilarity between each stimulus was identical. For unilateral, two models were defined. In the first, same-color stimuli are defined to have a dissimilarity of 0, while different-colored stimuli were defined to have a dissimilarity of 1. The second unilateral model was similar, but based on shape instead of color. The bilateral model is the mean of both unilateral models, corresponding to equal sensitivity to either dimension. For each subject, the correlation between empirical RDM and all model RDMs was computed, and the subject was assigned the model type with the highest correlation. A bootstrapping test was then performed within experiments, to test whether bilateral priors predicted reaching a bilateral stage in the main experiment. For 10.000 iterations, bootstrapped arena model assignments were randomly sampled with replacement from among all arena model assignments. Main task model order assignment was similarly bootstrapped. Then, for each iteration, the number of matches between arena and main task was computed, where a match was defined as either both models of a bootstrapped subject being bilateral, or neither model being bilateral. This was compared against the empirical number of matches. The setup and results of the arena analysis are displayed in Figure b-1. In the plus task, the observed number of matches was significantly greater than the number of bootstrapped matches in both experiments ($p \leq .023$). In the polar task, no differences were observed between bootstrapped and true matches ($p \geq .275$). These results are visualized in Figure b-1b. In conclusion, bilateral priors predicted subsequent bilateral generalization only in the grid task. A possible cause for this is that narrower transfer suffices for the grid task, as most bilateral arena arrangements were grids.

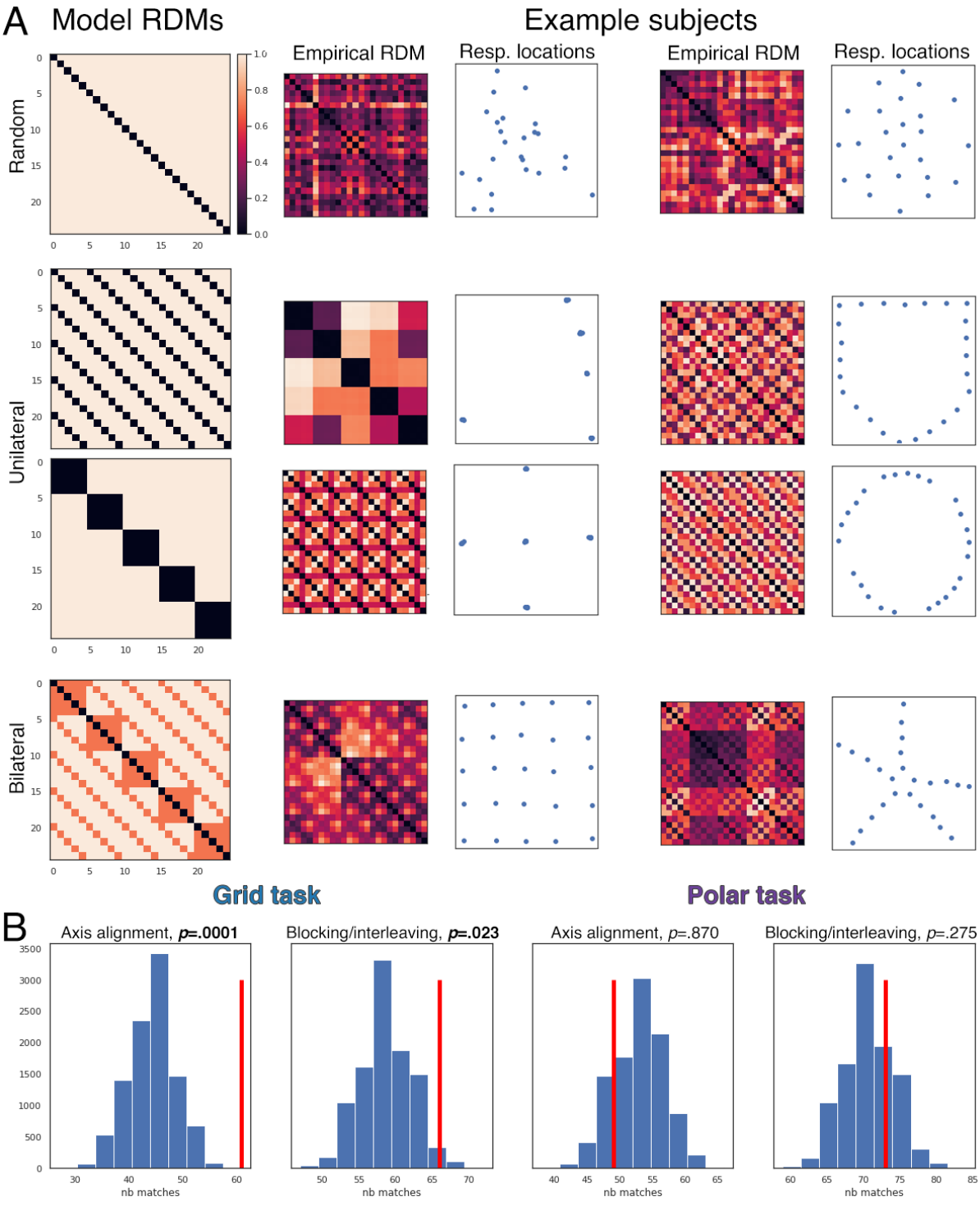

Figure c-1: Stimuli dissimilarity rating task analysis. (**A**) Left: model RDMs. The random model corresponds to sensitivity to neither dimension. Unilateral models correspond to sensitivity to only one dimension (shape or color). The bilateral model corresponds to equal sensitivity to either dimension. Right: Example subjects per assigned model type, with empirical RDM and corresponding response locations. (**B**): Results of a bootstrapping analysis, comparing the true

number of matches (red line) against a bootstrapped distribution of arena task and main task model assignments (blue), using 10.000 bootstrap iterations.

c.3 Cluster-based learning model

In the main text, we observed that the learning dynamics of human learners included a unilateral stage. What computational mechanism could underlie this? The model introduced here is a formalization of the idea that this process involves specialized clusters embodying elementary functions, along with a control mechanism that flexibly combines clusters to adapt to specific tasks. Our concrete implementation is a version of the mixture-of-experts setup (Jacobs et al., 1991), with a heuristic, input-based rule for cluster creation.

c.3.1 Methods

Definition. Let a task T be a set of inputs $X = \{x_1, x_2, \dots x_n\}$ with corresponding outputs $Y = \{y_1, y_2, \dots y_n\}$. To solve a task is to find a function that maps x_i onto y_i for all $i \mid i \in \mathbf{N}, i \leq n$. Our architecture consists of a controller Π and a set of modules $m \in M$. Both are fully linear without bias units, so we may interpret them as weight matrices. The dimensionality of Π is $(N_x \times N_M)$, the number of inputs (10 in our tasks) and modules, respectively. The output of the controller is $g = x\Pi$. The dimensionality of each module m is $(N_x \times N_y)$. For convenience, we concatenate all m into a single $(N_x \times N_y N_M)$ matrix M_{conc}, so that the experts' output $E = R(xM_{conc})$, where R is the reshape operation $A \in \mathbf{R}^{1 \times N_y N_M} \rightarrow A \in \mathbf{R}^{N_M \times N_y}$. The final model prediction is $\hat{y} = gE$, the vector-matrix product between gating signal and expert outputs.

Upon presentation of an input x_i, the model must determine whether to create a new cluster, or to predict an output with the existing modules. This is determined by the cluster definition CD_k of a module m_k. We rewrite the matrix Π as a set of column vectors, $\Pi = (a_1, a_2, \dots a_c)$. Now, $D_k = N_y \frac{a_k}{sum(a_k)}$. CD_k may be interpreted as a normalized measure of the inputs that maximally activate m_k. A standard prediction $\hat{y}$ is computed if $sum(abs(CD_k - x_i)) < \tau$ for all k, where τ is a threshold hyperparameter. If not, N_M is incremented by 1, a new randomly initialized module m_{N_M} is appended

to M, and $a_{N_M} = x_i$ is appended to Π, implying that initially, the input that maximally activates a module is the input that triggered its creation.

Both Π and Mwere optimized after each input x_i using the Adam optimization algorithm. The loss function was the mean squared error between $\hat{y}$ and y, i.e. $L_{main} = \frac{1}{N_Y}\sum_{i=1}^{N_Y}(\hat{y}_i - y_i)^2$. In addition, the term $L_{pos} = sum(ReLU(\Pi))$ was added to the loss, where ReLU is the elementwise operation $ReLU(x) = \max(0, x)$, incorporated to ensure controller weights were positive. The learning rate on Π was .05, and the learning rate on $M$was .01.

To provide intuition on how the currently introduced setup functions, Figure b-2 illustrates the effects of threshold and curriculum in a cluster learning setup.

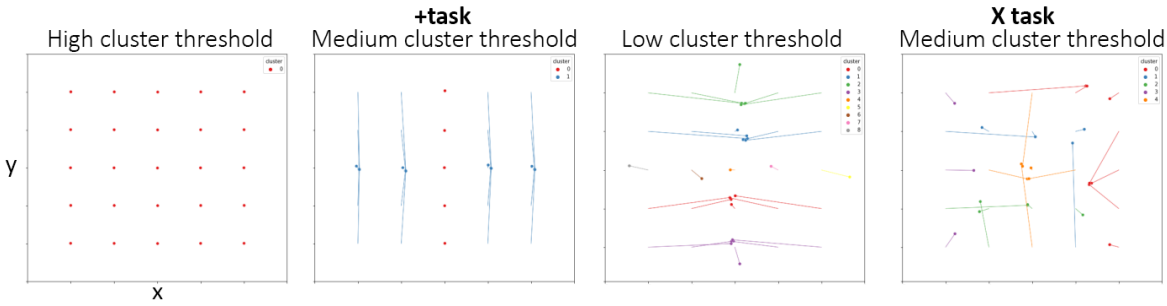

Figure c-2: Illustration of different clusterings depending on condition and threshold. The input space may be partitioned into different clusters depending on a subject's propensity to lump or split (threshold), the types of input and the order in which inputs are encountered. At one extreme, the same mechanism is assigned to all inputs (high threshold). At the other extreme, it is assumed that each input was generated through its own unique mechanism (high threshold). Presumably, the optimal structure for compositional inference lies somewhere in between.
Illustrated here are the results of a simplified version of the clustering model applied to our + and X tasks. In this version, an argmax was applied to *g* during both training and evaluation to highlight the contribution of the most strongly activated module for each location. The controller was updated with the delta rule $\Delta a = .3(a - x_i)$ for the activated $a \in \Pi$. Color indicates per-location cluster assignment. Points indicate model predictions, connected with a line to corresponding ground truth locations. Of note is that with an intermediate threshold, the model can learn a cluster structure matching the latent dimensionality in the axis-aligned setting, but not in the misaligned setting.

The model so far predicts why learners may display unilateral errors. A different question is which pressures give rise to the learning dynamics, which include qualitative changes in generalization behavior (for example, random to unilateral to bilateral). We consider that subjects likely prefer rules that involve fewer clusters, following Occam's razor and matching behavioral findings in which low-

dimensional category structures are more readily learned than high-dimensional structures (Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994). For this purpose, we introduce a *sparsity cost* based on Hoyer (2004):

$$L_{sparsity}(x) = 1 - \frac{\sqrt{n} - (\sum |x_i|) / \sqrt{\sum x_i^2}}{\sqrt{n} - 1} \quad \text{(Eq. b.1)}$$

with n the dimensionality of x. This is a scale-invariant metric, which is 0 when x is one-hot, 1 when all elements of x are equal, and 0 ≤ 1 for intermediate degrees of sparsity. Concretely, we add $L_{s1} = L_{sparsity}(g)$ to the loss term. Second, we assume that there is a preference for task-level solutions, i.e. using the same combination of modules regardless of the current input. The motivation for this is that top-down control is slow and physiologically costly in real organisms. As in a variety of learning tasks, reaction times decrease as a function of learning (Feeney, Howard & Howard, 2002; Schuck et al., 2015), this pressure was expected to particularly affect later learning stages. Concretely, we add $L_{s2} = 1 - \frac{t}{N_{trials}\, N_M} \sum_{i=1}^{N_M} L_{sparsity}\,(a_i)$ to the loss term, where t is the trial number. To summarize, for feedback trials, the loss is:

$$L_{feedback} = L_{main} + L_{pos} + L_{s1} + L_{s2} \quad \text{(Eq. b.2}$$

For generalization, L_{main} was not used, as the system did not have access to y. Sparsity costs could still be computed, as they only require x_i and internal states, so that::

$$L_{generalization} = L_{pos} + L_{s1} + L_{s2} \quad \text{(Eq. b.3)}$$

This concludes the model description. The model was evaluated on the grid task, on the + curriculum, the X curriculum, and an interleaved version of the + curriculum.

c.3.2 Results

Figure b-3 displays results of the full cluster model, with both sparsity pressures applied. Qualitatively, model behavior initially matches the random generalization pattern, then transitions through a unilateral stage, toward full bilateral generalization in the end. Moreover, the model reproduces sensitivity to

training curriculum observed in the behavioral experiment, with generalization errors (squared error) being lowest when training is blocked and axis-aligned. Interleaving and training on non-axis aligned locations both increased generalization errors.

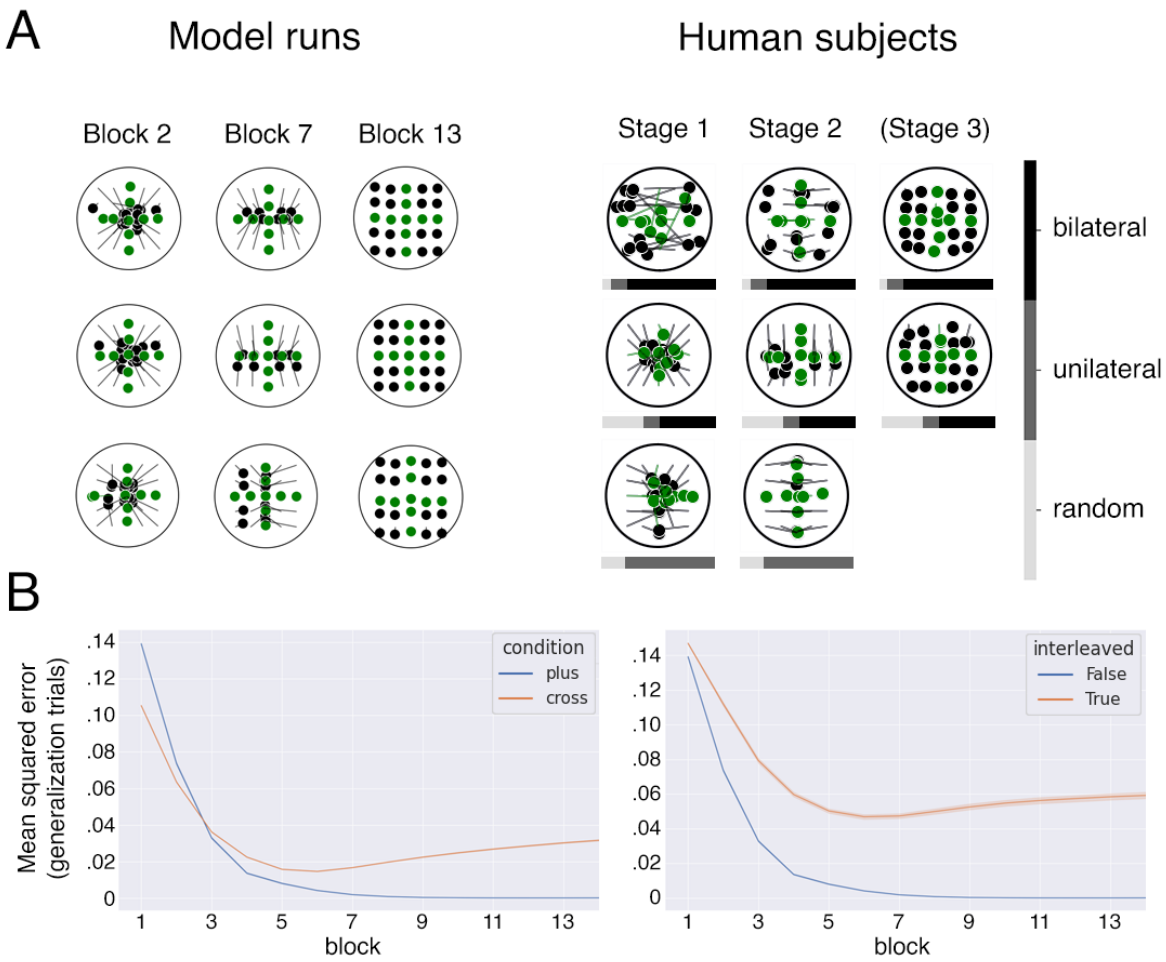

Figure c-3: Cluster model results. (**A**): Behavior in the + condition in the grid task. Training locations are displayed in green, while generalization is displayed in black. Each row is a single model run or subject, visualized at different time points (Left): Example model runs. Displayed are responses for blocks 2, 7 and 13. (Right): Example human subjects. Displayed are average responses per stage. Below each plot is a visualization of the model order fitted to that subject. (**B**): Mean squared error for generalization locations as a function of experimental block. Lines are model averages over 1000 runs. Error bars represent 95% CI. (Left): Effect of axis alignment. (Right): Effect of interleaving, applied to the + curriculum.

c.4 Meta-learning - recursive modular architecture

The cluster-based model introduced in the previous section provides a qualitative description of our behavioral results. One limitation is that with linear clusters, the model can only solve linear tasks. As humans are clearly capable of a wider variety of behavior, another important question is how humans learn appropriate elementary functions. It is not generally possible to learn any nonlinear function from sparse training examples. Instead, we posit that there are two possibilities. First, ecologically relevant modules may be to some extent hardcoded, i.e. architecture, priors or plasticity may be constrained to produce certain patterns of induction more readily than others. Second, general principles (such as rotation) may be learned by integrating across a wide variety of different experiences and tasks. We focus on the second possibility, which is more general and flexible. Indeed, in our polar task, it seems plausible that subjects come into the experiment pre-equipped with a notion of rotation, rather than inferring it from the 5 different angles presented in the experiment. Learning nonlinear functions brings with it the complication that different modules cannot be combined simply by taking a weighted sum. Instead, here we approach this problem with recursive computation (Caballero, Humphries, & Gurney, 2018; Chang et al., 2018), i.e. iteratively processing an internal representation until the desired composition of transformations is performed. For example, one could solve the polar task by first applying a scale transformation and then applying rotation (or vice-versa).

One desideratum is that the learner should be able to quickly learn a new task that has the same latent structure, but differs in superficial features. For example, a human learner who has mastered to polar task would likely be much quicker in subsequently learning a variation of the task in which the mapping from color levels to rule magnitudes was reshuffled. For this purpose, we include a task embedding network, which learns more quickly than the other elements and is wholly linear. Furthermore, we consider that elementary functions can be made more expressive by providing a *magnitude* by which to perform them. For example, it seems unnatural that different modules should be learned for different degrees of rotation, or that all rotations involve recursive application of a single core rotation amount.

c.4.1 Methods

Using the same task definition as before, given an input x_i, the model first converts this into an initial task representation through the linear embedding network E. The dimensionality of the task representation is a hyperparameter, the latent capacity. An internal representation of the output with dimensionality N_Y is also initialized. For our tasks, we initialized this as $[1,0]$, the point corresponding to $[\cos(0), \sin(0)]$ on the unit circle. The controller Π was a two-hidden layer neural network, with ReLUs nonlinearities in the hidden layers. All hidden layers had 64 nodes. The controller was optimized with Proximal Policy Optimization, for comparability with Chang et al (Chang et al., 2018). The controller receives the entire task representation and the current output representation as its input, and its function is to select the appropriate module, determined by the softmax distribution of its outputs. The number of modules is a hyperparameter. We used 3 in our simulations, allowing the model to in principle surpass the intrinsic task dimensionality of 2. The core functionality of a module is its task network, a 1-hidden layer MLP using the ReLU nonlinearity. Crucially, both the input and output of all task networks had dimensionality N_Y, enabling recursive processing. In addition, this implies that the task network has a local view, by which it does not have access to the task representation. This was done to prevent overfitting to the specific task. In addition, each module contained a hypernetwork, another 1-hidden layer MLP, which received inputs from the task representation condensed into a single scalar, and whose output had the same dimensionality as the hidden layer of its corresponding task network. The output of the hypernetwork was temporarily added to the weights of the hidden layer of its task network, modifying its function for that execution only. In principle, trainable components can be used to determine the mapping from task representation to hypernetwork input and how the task representation should be modified after a module call. For simplicity and interpretability, we instead used a rule whereby when module k was called, the hypernetwork received task representation node k as its input, and this node was then set to 0. In addition to these standard modules, a stop module was always included. When the controller selected the stop module, the trial ended, and the current output representation became the model's final answer $\hat{y}$. The cost function was simply the mean squared error, $loss = \frac{1}{N_Y}\sum_{i=1}^{N_Y}(\hat{y}_i - y_i)^2$, with a small penalty (.01) per step, to encourage shorter

solutions. The Adam algorithm was used for optimization, with a learning rate of $5 \cdot 10^{-2}$ for the embedding network, $5 \cdot 10^{-3}$ for the modules, and $5 \cdot 10^{-5}$ for the controller. The model trained on each task for 28800 steps, where each module call was a step. The embedding network was updated after every trial, while the modules and controller were updated every 4800 steps, for a total of 6 times per task. The relatively large plasticity of the embedding network is important, as this reduces overwriting of the controller function and module content in the initial stages of a task. The model is visualized in Figure b-4 and b-5.

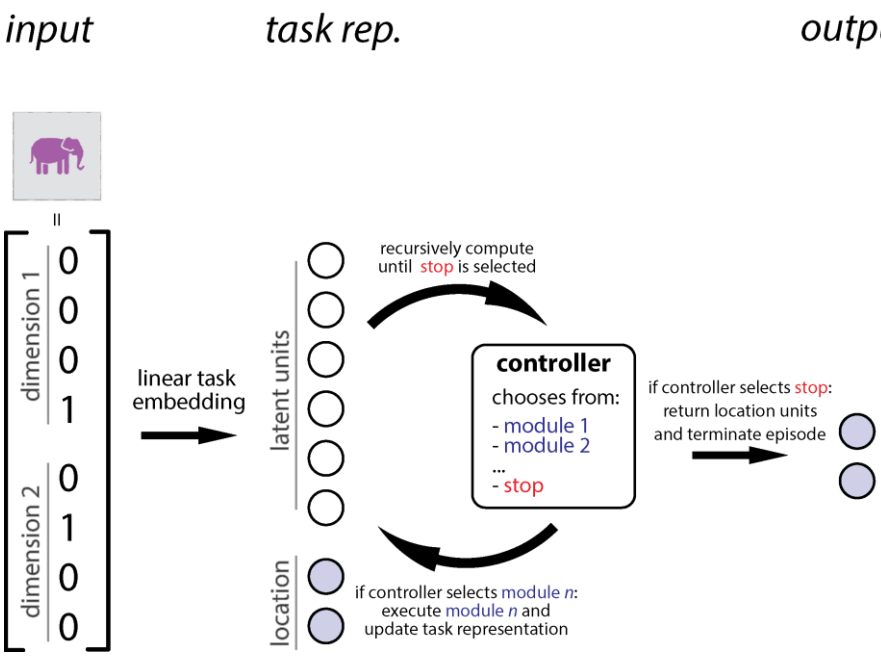

Figure c-4: Recursive modular architecture, rolled-up view. Physical stimulus dimensions are modeled as k-hot vectors. These are first converted into a task representation through an embedding network. Based on the latent representation, a controller iteratively selects modules to transform it by, until the stop module is selected, at which point the current output estimate is returned as the model's final output.

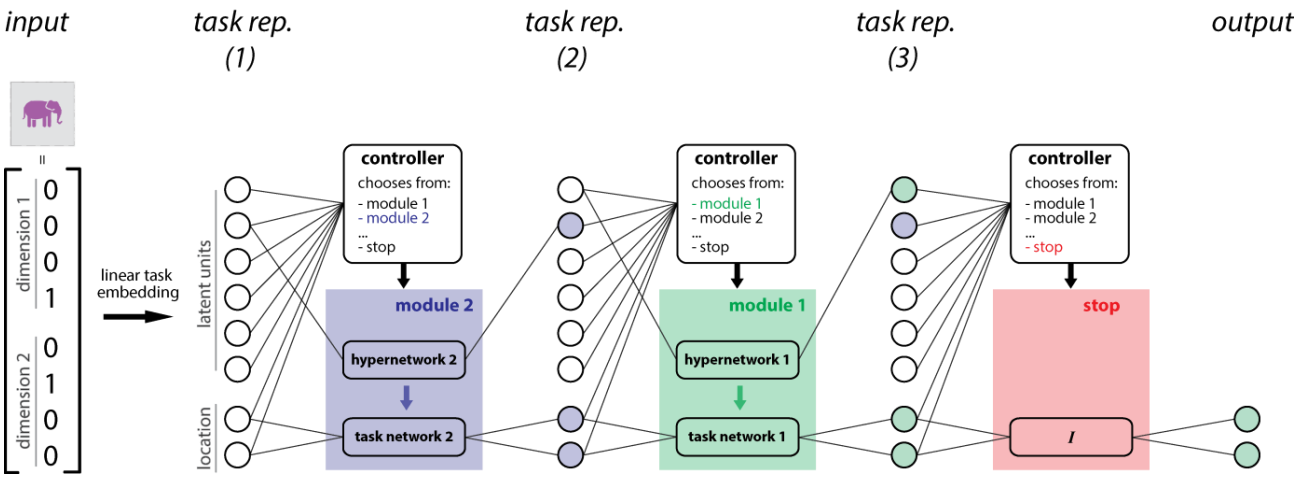

Figure c-5: Module structure and example rollout. Each module contained a task network with a local view, receiving the internal output estimate, but not the internal task representation as its inputs. Task networks were transiently modified by a corresponding hypernetwork, which received a single scalar as input, and modified the task network. An example rollout is displayed, in which two modules are called before the controller selects the stop module.

This model was used in a meta-learning setup, with the aim of gradually extracting the latent generative functions. Meta-learning consisted of training on randomly generated tasks, each consisting of 2 dimensions with 4 levels. Each dimension was randomly assigned to a rotation rule, scale rule, or was a dummy dimension with no function from among a curriculum-dependent set of eligible rules (Table S1). Rules were assigned to dimensions with replacement. The levels were assigned randomized quantities to perform the operation by, drawn from the interval $[-\pi, \pi]$ for rotation, and $[-3,3]$ for scale. Assuming a start at [0,1], the rotation rule was to rotate by the magnitude represented by each dimension level, while the scale rule was to multiply by the corresponding amount. When both dimensions indicated the scale same rule, the ground truth was to perform the corresponding transformation by the sum of the magnitudes indicated by either dimension. We investigated the effects of training on different task distributions, as indicated in Table b-1. Specifically, we considered the setting where the model first fully trains on one dimension, and then on the other (setup 1). We then tested whether the model was able to extract the second rule from training only on tasks which involved both rule 1 and 2 (setup 2). Finally, we tested if the model was able to create appropriate modules from training on the full-complexity task distribution immediately (setup 3). After meta-training in this fashion, the model was always evaluated on a version of the polar task, i.e. a task involving rule 1 and 2. During meta-learning, full feedback was provided for all locations. For the

transfer task at the end, feedback was provided only for a single ring and spoke, as in the human experiment.

	Task distribution 1	Number of tasks	Task distribution 2	Number of tasks
Setup 1	Rule 1, dummy	100	Rule 2, dummy	100
Setup 2	Rule 1, dummy	100	Rule 1 & 2 (without replacement)	100
Setup 3	Rule 1 & 2 (without replacement)	200	-	0

Table b-1: Meta learning curricula. In each setup, the model was first trained on task distribution 1 for 100 or 200 tasks. It was then trained on task distribution 2 for 100 or 0 tasks. Rule 1 and 2 were randomly assigned to the rotation and scale rules.

c.4.2 Results

In setup 1, training on the individual rules transferred to the pirate task. Without prior training on the composition of rule 1 and 2, the model reached a mean squared error of 0.003 in generalization. In both setup 1 and 2, the model succeeded at extracting modules which qualitatively matched the underlying rules of scale and rotation. The modules learned in setup 2 are visualized in Figure b-6a. In setup 3, the network instead relied on a single module solution, never calling the other modules. The representation appears to capture scale and a limited set of angles, visualized in Figure b-6b. In transfer to the pirate task, after training on setup 3, the model had a mean squared error of 1.74 in generalization.

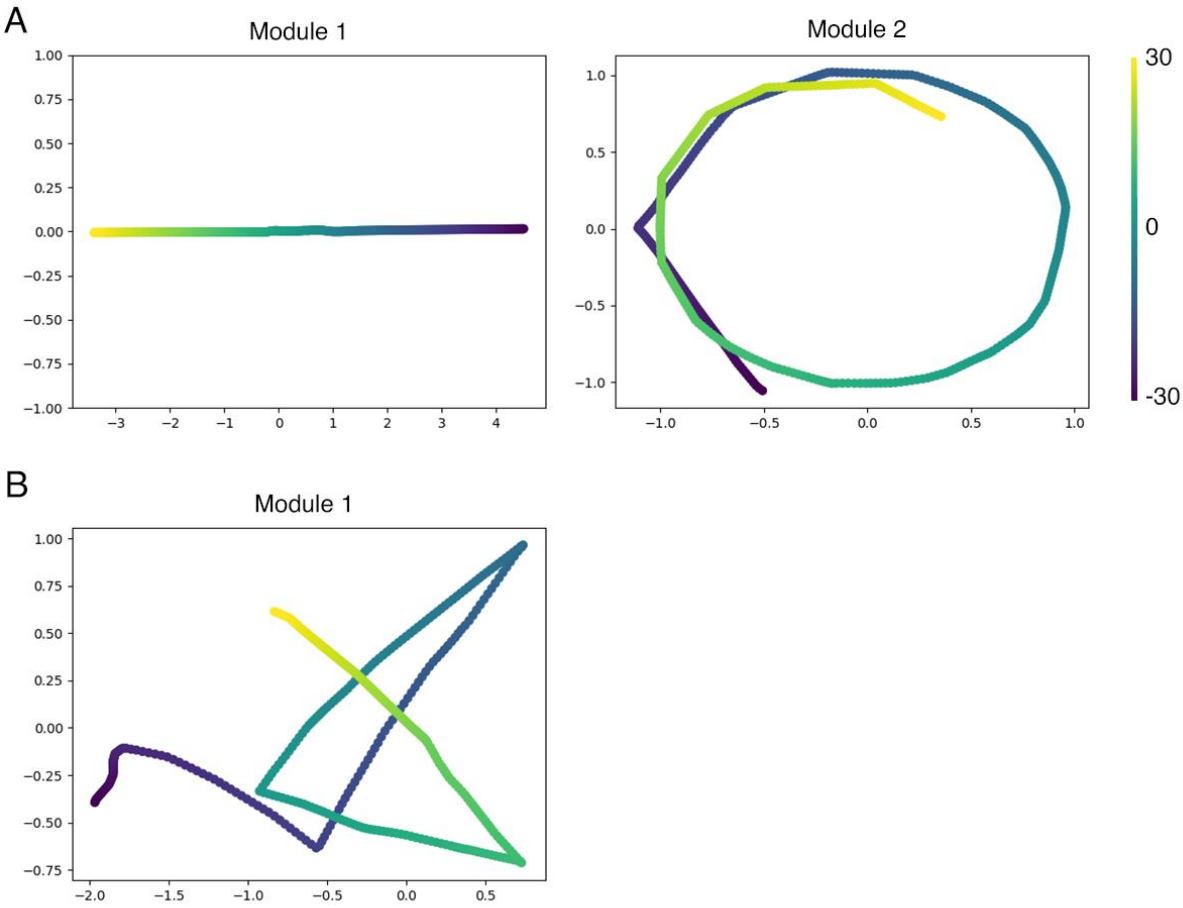

Figure c-6: Per-module latent dimension traversal. Each plot displays the output of a module receiving a (1,0) location input and a "magnitude", the scalar input of its hypernetwork, displayed from navy for low values to yellow for high values after training on task distribution 2. (**A**): Modules learned in setup 2. The network learned a bimodular solution corresponding to the generative ruleset. (**B**): Module learned in setup 3. The network learned a single-module solution for each task, with entangled rules.

c.4.3 Discussion

The simulations in this section concerned themselves with a modular architecture, with the aim of learning modules that match the generative structure of the environment. Specifically, we considered different training sequences in a meta-learning setting. The first setup trained on both rules in series, with a separate task set for either rule. This ensured that the training distribution did not feature tasks that required the composition of both rules. Despite this, the model obtained near-perfect generalization from sparse feedback on the composite task, highlighting how modular representation

may enable few-shot learning in compositional settings. The second setup was focused on distilling the second generative rule from composite tasks, when the first rule was already learned. An example of this would be learning a rotation module through tasks that use both scale and rotation on every trial, when one already knows how to use scale. Indeed, the model was able to learn modules matching the generative structure, even without training on the second rule in isolation. The third setup also challenged the network to infer appropriate modules from the composite setting, but now without first training on a single rule. This resulted in a failure mode, in which the network came to rely on single-module solutions overfitted to specific tasks, representing rotation and scale in an entangled fashion. When evaluated on the polar task with sparse feedback, the model failed to generalize to held-out locations.

Of particular interest is to consider these results in light of previously observed differences in the optimal training conditions of humans and artificial learners. When training in settings which require learners to integrate over multiple trials to discover within-context structure, human learners benefit from blocking contexts, experiencing greater interference when contexts are not clustered in time. By contrast, standard neural networks display the converse, performing best when training on the (interleaved) test distribution, and experiencing greater interference with blocked training due to catastrophic forgetting (Flesch et al., 2018). In this regard, the modular architecture displays several similarities to human learners, as both learn better when trained in a blocked fashion, and are capable of compositional generalization.

c.5 Additional model fitting visualizations

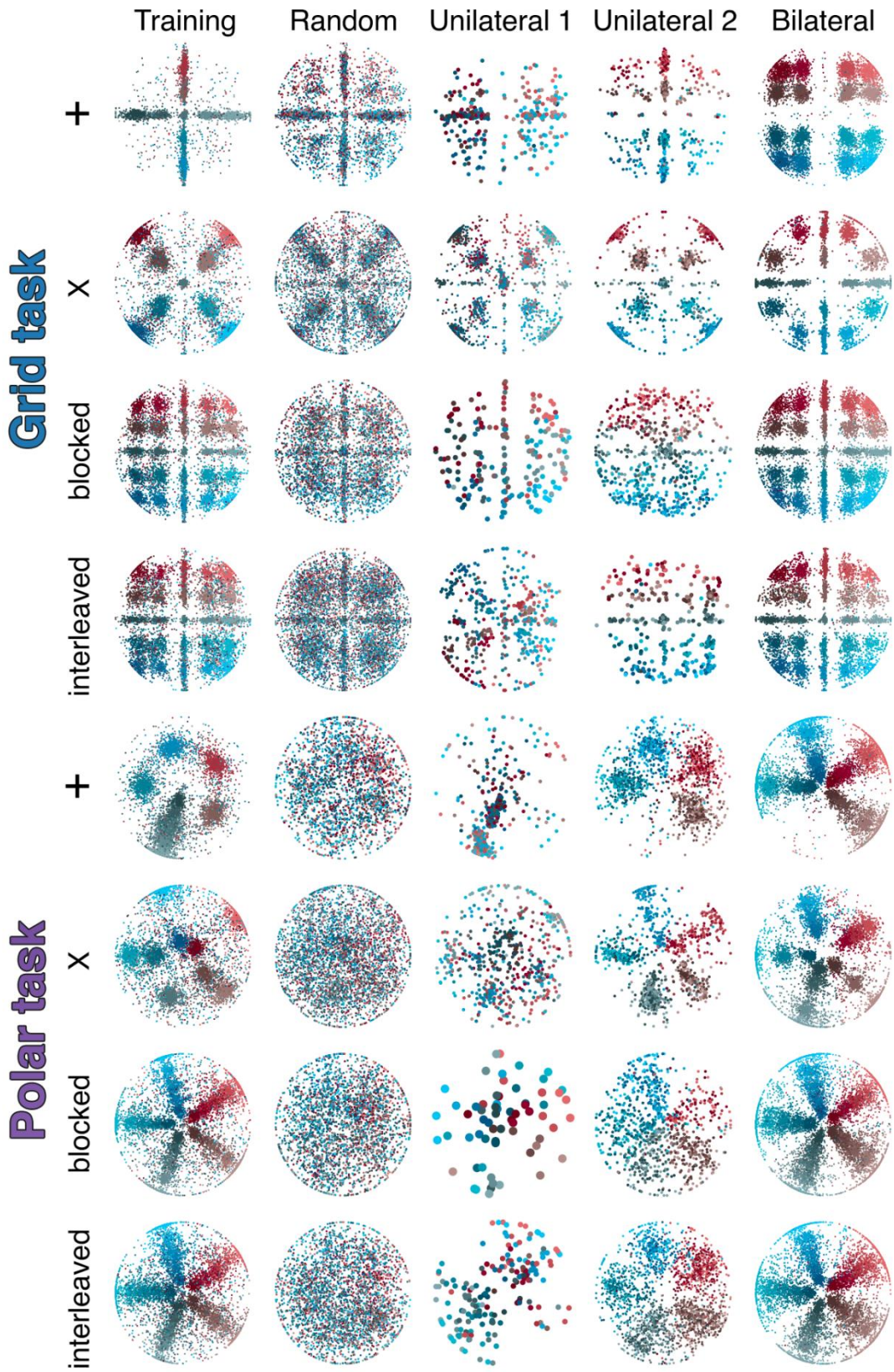

Figure c-7: Responses per model per condition: Visualized is the same type of plot as in figure 4-6b, but here separated by condition instead of task. The left column displays all feedback location responses. The remaining columns display generalization location responses, sorted by per-block model according to our model order fitting procedure. Hue indicates y/eccentricity dimension, while lightness indicates x/angle dimension of ground truth.

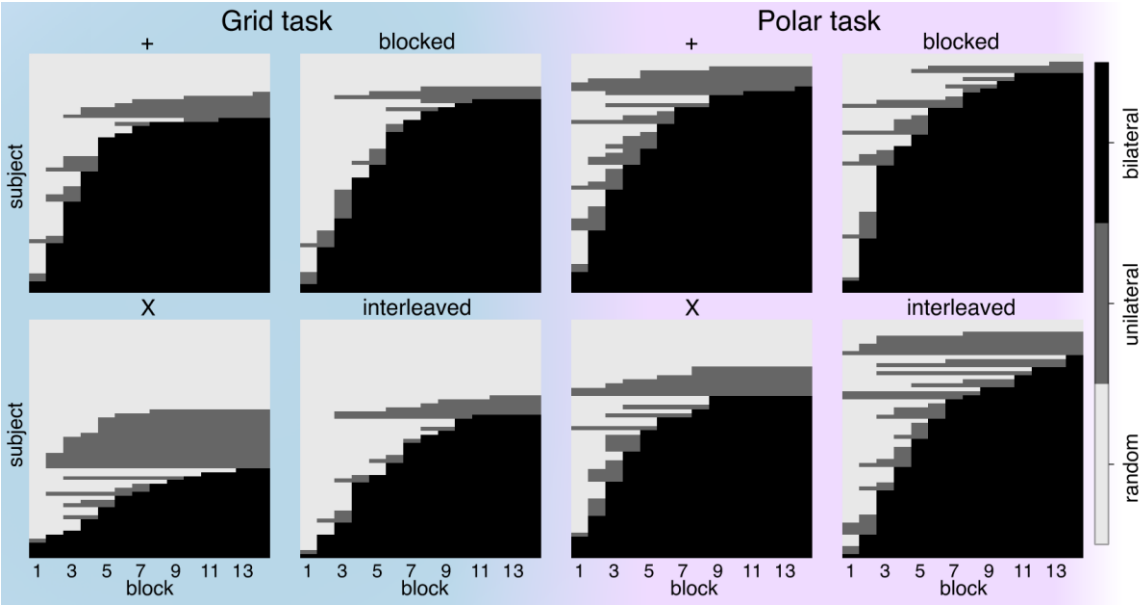

Figure c-8: Fitted model orders per subject. Each plot represents a condition, where each row is a subject, and each column is a block. Color indicates the model assigned to each block for each subject. White indicates the random model, grey the unilateral model and black the bilateral model.

Bibliography

Ahmad, S., & Tesauro, G. (1988). Scaling and generalization in neural networks: a case study. *Advances in Neural Information Processing Systems*, *1*, 160–168.

Akaishi, R., Kolling, N., Brown, J. W., & Rushworth, M. (2016). Neural mechanisms of credit assignment in a multicue environment. *Journal of Neuroscience*, *36*(4), 1096–1112.

Akrout, M., Wilson, C., Humphreys, P. C., Lillicrap, T., & Tweed, D. (2019). Deep Learning without Weight Transport. *Advances in Neural Information Processing Systems*. Retrieved from http://arxiv.org/abs/1904.05391

Allport, A., Styles, E. A., & Hsieh, S. (1994). 17 Shifting Intentional Set: Exploring the Dynamic Control of Tasks.

Ashby, Alfonso-Reese, L. A., Turken, U., & Waldron, E. M. (1998). *A Neuropsychological Theory of Multiple Systems in Category Learning*. *Psychological Review* (Vol. 105). Retrieved from https://psycnet.apa.org/journals/rev/105/3/442.html?uid=1998-04671-002

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*. Blackwell Publishing Inc. https://doi.org/10.1111/j.1749-6632.2010.05874.x

Ashby, & Maddox, W. (2005). Human category learning. *Annu. Rev. Psychol.* Retrieved from http://www.annualreviews.org/doi/abs/10.1146/annurev.psych.56.091103.070217

Atito, S., Awais, M., & Kittler, J. (2021). SiT: Self-supervised vIsion Transformer. *ArXiv Preprint ArXiv:2104.03602*.

Averell, L., & Heathcote, A. (2011). The form of the forgetting curve and the fate of memories. *Journal of Mathematical Psychology*, *55*(1), 25–35.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612.

Bechtel, W., & Abrahamsen, A. (1991). *Connectionism and the mind: An introduction to parallel processing in networks.* Basil Blackwell.

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., … Pouget, A. (2008). Probabilistic Population Codes for Bayesian Decision Making. *Neuron*, *60*(6), 1142–1152.

https://doi.org/10.1016/j.neuron.2008.09.021

Becker, S., & Hinton, G. E. (1992). Self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, *355*(6356), 161–163.

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*(2), 490–509.

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.

Belkin, M., Hsu, D., & Xu, J. (2020). Two models of double descent for weak features. *SIAM Journal on Mathematics of Data Science*, *2*(4), 1167–1180.

Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning. *Proceedings of the 26th Annual International Conference on Machine Learning*, 41–48. https://doi.org/10.1145/1553374.1553380

Bernardi, S., Benna, M. K., Rigotti, M., Munuera, J., Fusi, S., & Salzman, C. D. (2020). The geometry of abstraction in the hippocampus and prefrontal cortex. *Cell*, *183*(4), 954–967.

Berner, C., Brockman, G., Chan, B., Cheung, V., Dębiak, P., Dennison, C., … Hesse, C. (2019). Dota 2 with large scale deep reinforcement learning. *ArXiv Preprint ArXiv:1912.06680*.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, *94*(2), 115.

Binder, M. D., Hirokawa, N., & Windhorst, U. (2009). *Encyclopedia of neuroscience* (Vol. 3166). Springer Berlin, Germany.

Binz, M., Gershman, S. J., Schulz, E., & Endres, D. (2020). Heuristics From Bounded Meta-Learned Inference.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo‐planar MRI. *Magnetic Resonance in Medicine*, *34*(4), 537–541.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the Real World: Essays Illustrating*

Fundamental Contributions to Society, *2*(59–68).

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. *From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes*, *2*, 35–67.

Bowman, C. R., Iwashita, T., & Zeithamova, D. (2020). Tracking prototype and exemplar representations in the brain across learning. *ELife*, *9*, 1–47. https://doi.org/10.7554/eLife.59360

Bowman, C. R., & Zeithamova, D. (2018). Abstract memory representations in the ventromedial prefrontal cortex and hippocampus support concept generalization. *Journal of Neuroscience*, *38*(10), 2605–2614. https://doi.org/10.1523/JNEUROSCI.2811-17.2018

Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data* (pp. 93–104).

Bröder, A. (2000). Assessing the empirical validity of the" Take-the-best" heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1332.

Bröder, A., & Schiffer, S. (2003). Take The Best versus simultaneous feature matching: Probabilistic inferences from memory and effects of reprensentation format. *Journal of Experimental Psychology: General*, *132*(2), 277.

Brooks, D. N., Rosch, E., & Lloyd, B. B. (1978). Nonanalytic concept formation and memory for instances. *Cognition and Categorization*. Retrieved from https://philpapers.org/rec/BRONCF

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology*, *20*(4), 493–523.

Bryant, D. J. (2014). Strategy selection in cue-based decision making. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *68*(2), 97.

Busemeyer, J. R., Dewey, G. I., & Medin, D. L. (1984). Evaluation of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 638–648. https://doi.org/10.1037/0278-7393.10.4.638

Butts, D. A., & Goldman, M. S. (2006). Tuning Curves, Neuronal Variability, and Sensory Coding. *PLoS Biology*, *4*(4), e92. https://doi.org/10.1371/journal.pbio.0040092

Caballero, J. A., Humphries, M. D., & Gurney, K. N. (2018). A probabilistic, distributed, recursive mechanism for decision-making in the brain. *PLoS Computational Biology*, *14*(4), e1006033.

Cao, Y., Summerfield, C., & Saxe, A. (2020). Characterizing emergent representations in a space of candidate learning rules for deep networks. *Advances in Neural Information Processing Systems*, *33*.

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). Emerging Properties in Self-Supervised Vision Transformers. Retrieved from https://github.com/facebookresearch/dino

Carpenter, K. L., Wills, A. J., Benattayallah, A., & Milton, F. (2016). A Comparison of the neural correlates that underlie rule‐based and information‐integration category learning. *Human Brain Mapping*, *37*(10), 3557–3574.

Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition*, *41*(5), 671–682.

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory and Cognition*, *42*(3), 481–495. https://doi.org/10.3758/s13421-013-0371-0

Carvalho, P. F., & Goldstone, R. L. (2015, April 30). What you learn is more than what you see: What can sequencing effects tell us about inductive category learning? *Frontiers in Psychology*. Frontiers Research Foundation. https://doi.org/10.3389/fpsyg.2015.00505

Carvalho, P. F., & Goldstone, R. L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1699.

Challis, B. H. (1993). Spacing effects on cued-memory tests depend on level of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 389.

Chang, H. S., Miller, E., & McCallum, A. (2017). Active bias: Training more accurate neural networks by emphasizing high variance samples. *ArXiv*.

Chang, M. B., Gupta, A., Levine, S., & Griffiths, T. L. (2018). Automatically composing representation transformations as a means for generalization. *ArXiv Preprint ArXiv:1807.04640*.

Chen, Z., Ding, R., Chin, T.-W., & Marculescu, D. (2018). Understanding the impact of label granularity

on cnn-based image classification. In *2018 IEEE international conference on data mining workshops (ICDMW)* (pp. 895–904). IEEE.

Chew, B., Hauser, T. U., Papoutsi, M., Magerkurth, J., Dolan, R. J., & Rutledge, R. B. (2019). Endogenous fluctuations in the dopaminergic midbrain drive behavioral choice variability. *Proceedings of the National Academy of Sciences*, *116*(37), 18732–18737.

Chomsky, N. (2017). The language capacity: architecture and evolution. *Psychonomic Bulletin & Review*, *24*(1), 200–203.

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, *6*(1), 27755. https://doi.org/10.1038/srep27755

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, *2*(10), 406–416.

Coenen, A., Nelson, J. D., & Gureckis, T. M. (2019). Asking the right questions about the psychology of human inquiry: Nine open challenges. *Psychonomic Bulletin & Review*, *26*(5), 1548–1587.

Cordes, D., Haughton, V. M., Arfanakis, K., Wendt, G. J., Turski, P. A., Moritz, C. H., … Meyerand, M. E. (2000). Mapping functionally related regions of brain with functional connectivity MR imaging. *American Journal of Neuroradiology*, *21*(9), 1636–1644.

Cormier, S. M. (2014). *Basic processes of learning, cognition, and motivation*. Psychology Press.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In *Simple heuristics that make us smart* (pp. 97–118). Oxford University Press.

Daunizeau, J., Adam, V., & Rigoux, L. (2014). VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. *PLoS Computational Biology*, *10*(1). https://doi.org/10.1371/journal.pcbi.1003441

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*(7), 571.

del Ojo Balaguer, J. (2018). Decision-making with hierarchical representations in humans. University of Oxford.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern*

recognition (pp. 248–255). Ieee.

Detorakis, G., Bartley, T., & Neftci, E. (2019). Contrastive Hebbian learning with random feedback weights. *Neural Networks*, *114*, 1–14.

Detterman, D. K. (1993). The case for the prosecution: Transfer as an epiphenomenon.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv Preprint ArXiv:1810.04805*.

Dinstein, I., Heeger, D. J., Lorenzi, L., Minshew, N. J., Malach, R., & Behrmann, M. (2012). Unreliable evoked responses in autism. *Neuron*, *75*(6), 981–991.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., … Gelly, S. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *ArXiv Preprint ArXiv:2010.11929*.

Duan, C. A., Erlich, J. C., & Brody, C. D. (2015). Requirement of prefrontal and midbrain regions for rapid executive control of behavior in the rat. *Neuron*, *86*(6), 1491–1503.

Duncker, L., Driscoll, L., Shenoy, K. V, Sahani, M., & Sussillo, D. (2020). Organizing recurrent network dynamics by task-computation to enable continual learning. *Advances in Neural Information Processing Systems*, *33*.

Dunsmoor, J. E., & Paz, R. (2015). Fear generalization and anxiety: behavioral and neural mechanisms. *Biological Psychiatry*, *78*(5), 336–343.

Dymond, S., Dunsmoor, J. E., Vervliet, B., Roche, B., & Hermans, D. (2015). Fear generalization in humans: systematic review and implications for anxiety disorder research. *Behavior Therapy*, *46*(5), 561–582.

Earle, S., Edwards, M., Khalifa, A., Bontrager, P., & Togelius, J. (2021). *Learning Controllable Content Generators*. Retrieved from http://arxiv.org/abs/2105.02993

Ebbinghaus, H. (1885). *Über das gedächtnis: untersuchungen zur experimentellen psychologie*. Duncker & Humblot.

Elio, R., & Anderson, J. R. (1984). *The effects of information order and learning mode on schema abstraction*. *Memory & Cognition* (Vol. 12).

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*(1), 71–99.

Erhan, D., Bengio, Y., Courville, A., Manzagol, P. A., Vincent, P., & Bengio, S. (2010). Why does unsupervised pre-training help deep learning? *Journal of Machine Learning Research*, *11*, 625–660. https://doi.org/10.1145/1756006.1756025

Farajtabar, M., Azizan, N., Mott, A., & Li, A. (2020). Orthogonal gradient descent for continual learning. In *International Conference on Artificial Intelligence and Statistics* (pp. 3762–3773). PMLR.

Feeney, J. J., Howard Jr, J. H., & Howard, D. V. (2002). Implicit learning of higher order sequences in middle age. *Psychology and Aging*, *17*(2), 351.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2009). Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(9), 1627–1645.

Ferguson, G. A. (1956). On transfer and the abilities of man. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *10*(3), 121.

Fernando, C., Banarse, D., Blundell, C., Zwols, Y., Ha, D., Rusu, A. A., … Wierstra, D. (2017). PathNet: Evolution Channels Gradient Descent in Super Neural Networks. *ArXiv Preprint ArXiv:* Retrieved from https://arxiv.org/abs/1701.08734

Fillingham, J. K., Hodgson, C., Sage, K., & Lambon Ralph, M. A. (2003). The application of errorless learning to aphasic disorders: A review of theory and practice. *Neuropsychological Rehabilitation*, *13*(3), 337–363.

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2010.01.003

Flesch, T., Balaguer, J., Dekker, R., Nili, H., & Summerfield, C. (2018). Comparing continual task learning in minds and machines. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(44), E10313–E10322. https://doi.org/10.1073/pnas.1800755115

Flesch, T., Juechems, K., Dumbalska, T., Saxe, A., & Summerfield, C. (2021). Rich and lazy learning of task representations in brains and neural networks. *BioRxiv*.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1–2), 3–71.

Fodor, J., & Lepore, E. (1996). The red herring and the pet fish: Why concepts still can't be prototypes.

Cognition, *58*(2), 253–270. https://doi.org/10.1016/0010-0277(95)00694-X

Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience*, *8*(9), 700–711.

Fox, M. D., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2007). Intrinsic fluctuations within cortical systems account for intertrial variability in human behavior. *Neuron*, *56*(1), 171–184.

Franklin, N. T., & Frank, M. J. (2020). Generalizing to generalize: humans flexibly switch between compositional and conjunctive structures during reinforcement learning. *PLoS Computational Biology*, *16*(4), e1007720.

Freedman, D. J., & Assad, J. A. (2016, July 8). Neuronal Mechanisms of Visual Categorization: An Abstract View on Decision Making. *Annual Review of Neuroscience*. Annual Reviews Inc. https://doi.org/10.1146/annurev-neuro-071714-033919

French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, *3*(4), 128–135.

Frensch, P. A., & Rünger, D. (2003). Implicit learning. *Current Directions in Psychological Science*, *12*(1), 13–18.

Fuentes, A. (2017). *The International Encyclopedia of Primatology, 3 Volume Set*. John Wiley & Sons.

Garnelo, M., Arulkumaran, K., & Shanahan, M. (2016). Towards deep symbolic reinforcement learning. *ArXiv Preprint ArXiv:1609.05518*.

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: carving nature at its joints. *Current Opinion in Neurobiology*, *20*(2), 251–256.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*(3), 306–355.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, *103*(4), 650.

Gigerenzer, G., & Selten, R. (2002). *Bounded rationality: The adaptive toolbox*. MIT press.

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In *Simple heuristics that make us smart* (pp. 3–34). Oxford University Press.

Giguère, G., & Love, B. C. (2013). Limits in decision making arise from limits in memory retrieval. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(19),

7613–7618. https://doi.org/10.1073/pnas.1219674110

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440–1448).

Gkioxari, G., Hariharan, B., Girshick, R., & Malik, J. (2014). R-cnns for pose estimation and action detection. *ArXiv Preprint ArXiv:1406.5212*.

Glaser, R. (1989). *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. Psychology Press.

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256). JMLR Workshop and Conference Proceedings.

Gluck, M. A., Shohamy, D., & Myers, C. (2002). How do people solve the “weather prediction” task?: Individual variability in strategies for probabilistic category learning. *Learning & Memory*, *9*(6), 408–418.

Golan, T., Raju, P. C., & Kriegeskorte, N. (2020). Controversial stimuli: Pitting neural networks against each other as models of human cognition. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(47), 29330–29337. https://doi.org/10.1073/PNAS.1912334117

Golkar, S., Lipshutz, D., Bahroun, Y., Sengupta, A., & Chklovskii, D. (2020). A simple normative network approximates local non-Hebbian learning in the cortex. *Advances in Neural Information Processing Systems*, *33*.

Goode, S., & Magill, R. A. (1986). Contextual interference effects in learning three badminton serves. *Research Quarterly for Exercise and Sport*, *57*(4), 308–314.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … Bengio, Y. (2014). Generative adversarial networks. *ArXiv Preprint ArXiv:1406.2661*.

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. *ArXiv Preprint ArXiv:1412.6572*.

Gould, I. C., Nobre, A. C., Wyart, V., & Rushworth, M. F. S. (2012). Effects of decision variables and intraparietal stimulation on sensorimotor oscillatory activity in the human brain. *Journal of Neuroscience*, *32*(40), 13805–13818.

Graves, A., Bellemare, M. G., Menick, J., Munos, R., & Kavukcuoglu, K. (2017). Automated curriculum

learning for neural networks. In *international conference on machine learning* (pp. 1311–1320). PMLR.

Güçlü, U., & van Gerven, M. A. J. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *35*(27), 10005–10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015

Gureckis, T. M., & Markant, D. B. (2012). Self-directed learning: A cognitive and computational perspective. *Perspectives on Psychological Science*, *7*(5), 464–481.

Hacohen, G., & Weinshall, D. (2019). On the power of curriculum learning in training deep networks. *36th International Conference on Machine Learning, ICML 2019*, *2019-June*, 4483–4496. Retrieved from http://arxiv.org/abs/1904.03626

Haidet, P., Morgan, R. O., O'malley, K., Moran, B. J., & Richards, B. F. (2004). A controlled trial of active versus passive learning strategies in a large group setting. *Advances in Health Sciences Education*, *9*(1), 15–27.

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2005). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. https://doi.org/10.1016/j.biopsycho.2005.04.001

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, *14*(1), 31–38.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026–1034).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Hebb, D. O. (1949). The organization of behavior; a neuropsycholocigal theory. *A Wiley Book in Clinical Psychology*, *62*, 78.

Herding, J., Ludwig, S., von Lautz, A., Spitzer, B., & Blankenburg, F. (2019). Centro-parietal EEG potentials index subjective evidence and confidence during perceptual decision making. *NeuroImage*, *201*, 116011. https://doi.org/10.1016/j.neuroimage.2019.116011

Herzog, M. H., Aberg, K. C., Frémaux, N., Gerstner, W., & Sprekeler, H. (2012). Perceptual learning, roving and the unsupervised bias. *Vision Research*, *61*, 95–99. https://doi.org/10.1016/j.visres.2011.11.001

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., … Lerchner, A. (2017). B -VAE: Learning basic visual concepts with a constrained variational framework. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*.

Higgins, I., Sonnerat, N., Matthey, L., Pal, A., Burgess, C. P., Bosnjak, M., … Lerchner, A. (2017). Scan: Learning hierarchical compositional visual concepts. *ArXiv Preprint ArXiv:1707.03389*.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *ArXiv Preprint ArXiv:1207.0580*.

Hintzman, D. L., & Ludlam, G. (1980). Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. *Memory & Cognition*, *8*(4), 378–382. https://doi.org/10.3758/BF03198278

Hoffrage, U., Garcia-Retamero, R., & Czienskowski, U. (2005). The robustness of the Take The Best Configural heuristic in linearly and nonlinearly separable environments. In *Proceedings of the 27th annual conference of the cognitive science society* (pp. 971–976). Citeseer.

Hogan, M. J., Carolan, L., Roche, R. A. P., Dockree, P. M., Kaiser, J., Bunting, B. P., … Lawlor, B. A. (2006). Electrophysiological and information processing variability predicts memory decrements associated with normal age-related cognitive decline and Alzheimer's disease (AD). *Brain Research*, *1119*(1), 215–226.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709. https://doi.org/10.1037/0033-295X.109.4.679

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2003). Errors in reward prediction are reflected in the event-related brain potential. *NeuroReport*, *14*(18), 241–281. https://doi.org/10.1097/00001756-200312190-00037

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and*

Memory, *7*(6), 418–439. https://doi.org/10.1037/0278-7393.7.6.418

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, *79*(8), 2554–2558.

Hornsby, A. N., & Love, B. C. (2014). Improved Classification of Mammograms Following Idealized Training. *Journal of Applied Research in Memory and Cognition*, *3*(2), 72–76. https://doi.org/10.1016/j.jarmac.2014.04.009

Houghton, J., & Sheehan, P. (2000). *A primer on the knowledge economy*. Victoria University.

Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, *5*(9).

Hu, M., & Nosofsky, R. M. (2021). Exemplar-model account of categorization and recognition when training instances never repeat. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001008

Huang, Y., & Yu, R. (2014). The feedback-related negativity reflects “more or less” prediction error in appetitive and aversive conditions. *Frontiers in Neuroscience*, (8 MAY). https://doi.org/10.3389/fnins.2014.00108

Hupkes, D., Dankers, V., Mul, M., & Bruni, E. (2019). The compositionality of neural networks: integrating symbolism and connectionism. *ArXiv Preprint ArXiv:1908.08351*.

Hupkes, D., Dankers, V., Mul, M., & Bruni, E. (2020). Compositionality decomposed: how do neural networks generalise? *Journal of Artificial Intelligence Research*, *67*, 757–795.

Ilgen, D. R., & Moore, C. F. (1987). Types and choices of performance feedback. *Journal of Applied Psychology*, *72*(3), 401.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation*, *3*(1), 79–87.

Jain, R., Isaksen, A., Holmgård, C., & Togelius, J. (2016). Autoencoders for level generation, repair, and recognition. In *Proceedings of the ICCC Workshop on Computational Creativity and Games* (p. 9).

Jongman, A., & Wade, T. (2007). Acoustic variability and perceptual learning. *Language Experience in Second Language Speech Learning*, 135–150.

Jordan, M. I., & Jacobs, R. A. (1992). Hierarchies of adaptive experts. In *Advances in neural*

information processing systems (pp. 985–992).

Kahana, M. J., & Adler, M. (2017). Note on the power law of forgetting. *BioRxiv*, 173765.

Kang, S. H. K., & Pashler, H. (2012). Learning Painting Styles: Spacing is Advantageous when it Promotes Discriminative Contrast. *Applied Cognitive Psychology*, *26*(1), 97–103. https://doi.org/10.1002/acp.1801

Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. *ArXiv Preprint ArXiv:1710.10196*.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8110–8119).

Kelly, S. P., & O'Connell, R. G. (2015). The neural processes underlying perceptual decision making in humans: Recent progress and future directions. *Journal of Physiology Paris*. https://doi.org/10.1016/j.jphysparis.2014.08.003

Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *The Modern Language Journal*, *75*(3), 305–313.

Kerr, R., & Booth, B. (1978). Specific and varied practice of motor skill. *Perceptual and Motor Skills*, *46*(2), 395–401.

Khalifa, A., Bontrager, P., Earle, S., & Togelius, J. (2020). Pcgrl: Procedural content generation via reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* (Vol. 16, pp. 95–101).

Khan, F., Mutlu, B., & Zhu, X. (2011). How do humans teach: On curriculum learning and teaching dimension. *Advances in Neural Information Processing Systems*, 1449–1457. Retrieved from http://papers.nips.cc/paper/4466-how-do-humans-teach-on-curriculum-learning-and-teaching-dimension

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., … Grabska-Barwinska, A. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, *114*(13), 3521–3526.

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, *1*(2), 106–120.

Konda, V. R., & Tsitsiklis, J. N. (2000). Actor-critic algorithms. In *Advances in neural information processing systems* (pp. 1008–1014). Citeseer.

Kornell, N., & Bjork, R. A. (2006). *Learning Concepts and Categories Is Spacing the "'Enemy of Induction'"?* Hintzman.

Krizhevsky, A. (2009). Learning Multiple Layers of Features from Tiny Images. Retrieved from https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet Classification with Deep Convolutional Neural Networks*. Retrieved from http://code.google.com/p/cuda-convnet/

Krueger, K. A., & Dayan, P. (2009). Flexible shaping: How learning in small steps helps. *Cognition*, *110*(3), 380–394. https://doi.org/10.1016/j.cognition.2008.11.014

Kruschke, J. K. (2005). Learning involves attention. *Connectionist Models in Cognitive Psychology*, *51*, 113–140.

Kumar, M. P., Packer, B., & Koller, D. (2010). Self-Paced Learning for Latent Variable Models. In *NIPS* (Vol. 1, p. 2).

Kurzawa, N., Summerfield, C., & Bogacz, R. (2017). Neural circuits trained with standard reinforcement learning can accumulate probabilistic information during decision making. *Neural Computation*, *29*(2), 368–393.

Lake, B., & Baroni, M. (2018). Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks. In *International Conference on Machine Learning* (pp. 2873–2882). PMLR.

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*.

Lampinen, A. K., & McClelland, J. L. (2020). Transforming task representations to perform novel tasks. *Proceedings of the National Academy of Sciences*, *117*(52), 32970–32981.

Lee, E. S., MacGregor, J. N., Bavelas, A., Mirlin, L., Lam, N., & Morrison, I. (1988). The Effects of Error Transformations on Classification Performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 66–74. https://doi.org/10.1037/0278-7393.14.1.66

Leshno, M., Lin, V. Y., Pinkus, A., & Schocken, S. (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, *6*(6), 861–867. https://doi.org/10.1016/S0893-6080(05)80131-5

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766.

Libby, A., & Buschman, T. J. (2021). Rotational dynamics reduce interference between sensory and memory representations. *Nature Neuroscience*, 1–12.

Lillicrap, T. P., Cownden, D., Tweed, D. B., & Akerman, C. J. (2016). Random synaptic feedback weights support error backpropagation for deep learning. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms13276

Lu, Z.-L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research*, *38*(9), 1183–1198.

Ludden, D., & Gupta, P. (2000). Zen in the art of language acquisition: Statistical learning and the less is more hypothesis. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 22).

Luyckx, F., Nili, H., Spitzer, B., & Summerfield, C. (2018). Neural structure mapping in human probabilistic reward learning. *BioRxiv*, 366757. https://doi.org/10.1101/366757

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. https://doi.org/10.1016/j.cub.2013.08.035

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 650.

Maddox, W. T., Ashby, F. G., Ing, A. D., & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory & Cognition*, *32*(4), 582–591.

Mammarella, N., Avons, S. E., & Russo, R. (2004). A short‐term perceptual priming account of spacing effects in explicit cued‐memory tasks for unfamiliar stimuli. *European Journal of*

Cognitive Psychology, *16*(3), 387–402.

Mammarella, N., Russo, R., & Avons, S. E. (2002). Spacing effects in cued-memory tasks for unfamiliar faces and nonwords. *Memory & Cognition*, *30*(8), 1238–1251.

Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Vintage.

Marcus, G. F. (2003). *The algebraic mind: Integrating connectionism and cognitive science*. MIT press.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Markant, D. B., & Gureckis, T. M. (2014). Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, *143*(1), 94.

Matiisen, T., Oliver, A., Cohen, T., & Schulman, J. (2019). Teacher–student curriculum learning. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(9), 3732–3740.

McCarthy, J. (1987). Generality in artificial intelligence. *Communications of the ACM*, *30*(12), 1030–1035.

McClelland, J. L., Rumelhart, D. E., & Group, P. D. P. R. (1986). *Parallel distributed processing* (Vol. 2). MIT press Cambridge, MA.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133.

Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(3), 333–352. https://doi.org/10.1037/0278-7393.10.3.333

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207

Meng, D., Zhao, Q., & Jiang, L. (2015). What objective does self-paced learning indeed optimize? *ArXiv Preprint ArXiv:1511.06049*.

Michel, N., Cater III, J. J., & Varela, O. (2009). Active versus passive teaching styles: An empirical study of student learning outcomes. *Human Resource Development Quarterly*, *20*(4), 397–418.

Milne, E. (2011). Increased intra-participant variability in children with autistic spectrum disorders: evidence from single-trial analysis of evoked EEG. *Frontiers in Psychology*, *2*, 51.

Milton, F., & Pothos, E. M. (2011). Category structure and the two learning systems of COVIS.

European Journal of Neuroscience, *34*(8), 1326–1336.

Minsky, M., & Papert, S. A. (1969). *Perceptrons: An introduction to computational geometry*. MIT press.

Misra, I., Shrivastava, A., Gupta, A., & Hebert, M. (2016). Cross-stitch networks for multi-task learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3994–4003).

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *ArXiv Preprint ArXiv:1312.5602*.

Montagu, A. (1961). Neonatal and Infant Immaturity in Man. *JAMA: The Journal of the American Medical Association*, *178*(1), 56–57. https://doi.org/10.1001/jama.1961.73040400014011

Morerio, P., Cavazza, J., Volpi, R., Vidal, R., & Murino, V. (2017). Curriculum dropout. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3544–3552).

Nagarajan, V., Andreassen, A., & Neyshabur, B. (2020). Understanding the failure modes of out-of-distribution generalization. *ArXiv Preprint ArXiv:2010.15775*.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*(1), 11–28.

Nili, H., Wingfield, C., Walther, A., & Su, L. (2014). A toolbox for representational similarity analysis. *PLoS Comput*. Retrieved from http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003553

Nomura, E. M., Maddox, W. T., Filoteo, J. V, Ing, A. D., Gitelman, D. R., Parrish, T. B., … Reber, P. J. (2007). Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex*, *17*(1), 37–43. https://doi.org/10.1093/cercor/bhj122

Nosofsky, R. M. (1988). Exemplar-Based Accounts of Relations Between Classification, Recognition, and Typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700–708. https://doi.org/10.1037/0278-7393.14.4.700

Nosofsky, R. M. (2020). Contrasting Exemplar and Prototype Models in a Natural-Science Category Domain, 641–647.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing modes of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition*, *22*(3), 352–369.

Nosofsky, R. M., Little, D. R., & James, T. W. (2012). Activation in the neural network responsible for

categorization and recognition reflects parameter changes. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(1), 333–338. https://doi.org/10.1073/pnas.1111304109

Novak, R., Bahri, Y., Abolafia, D. A., Pennington, J., & Sohl-Dickstein, J. (2018). Sensitivity and generalization in neural networks: an empirical study. *ArXiv Preprint ArXiv:1802.08760*.

Nye, M. I., Solar-Lezama, A., Tenenbaum, J. B., & Lake, B. M. (2020). Learning compositional rules via neural program synthesis. *ArXiv Preprint ArXiv:2003.05562*.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), 1729–1735. https://doi.org/10.1038/nn.3248

Park, W. J., Schauder, K. B., Zhang, R., Bennetto, L., & Tadin, D. (2017). High internal noise and poor external noise filtering characterize perception in autism spectrum disorder. *Scientific Reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-17676-5

Parpart, P., Jones, M., & Love, B. C. (2018). Heuristics as Bayesian inference under extreme priors. *Cognitive Psychology*, *102*, 127–144.

Pashler, H., & Mozer, M. (2013). When does fading enhance perceptual category learning? *Journal of Experimental Psychology:* Retrieved from http://psycnet.apa.org/journals/xlm/39/4/1162/

Passingham, R. E., & Wise, S. P. (2012). *The neurobiology of the prefrontal cortex: anatomy, evolution, and the origin of insight*. Oxford University Press.

Patil, K. R., & Love, B. C. (2014). Optimal Teaching for Limited-Capacity Human Learners. *Nips*, 1–9.

Pavlik, P. I., & Anderson, J. R. (2008). Using a Model to Compute the Optimal Schedule of Practice. *Psychological Association*, *14*(2), 101–117. https://doi.org/10.1037/1076-898X.14.2.101

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, *31*(2). https://doi.org/10.1017/S0140525X08003543

Peterson, J. C., Abbott, J. T., & Griffiths, T. L. (2017). Evaluating (and improving) the correspondence between deep neural networks and human representations. Retrieved from http://arxiv.org/abs/1706.02417

Pfabigan, D. M., Alexopoulos, J., Bauer, H., & Sailer, U. (2011). Manipulation of feedback expectancy

and valence induces negative and positive reward prediction error signals manifest in event-related brain potentials. *Psychophysiology*, *48*(5), 656–664. https://doi.org/10.1111/j.1469-8986.2010.01136.x

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*(1–2), 73–193.

Pinto, N., Doukhan, D., DiCarlo, J. J., Cox, D. D., Fukushima, K., Bishop, C., … Eberhart, R. (2009). A High-Throughput Screening Approach to Discovering Good Forms of Biologically Inspired Visual Representation. *PLoS Computational Biology*, *5*(11), e1000579. https://doi.org/10.1371/journal.pcbi.1000579

Poletiek, F. H., Conway, C. M., Ellefson, M. R., Lai, J., Bocanegra, B. R., & Christiansen, M. H. (2018). Under What Conditions Can Recursion Be Learned? Effects of Starting Small in Artificial Grammar Learning of Center‐Embedded Structure. *Cognitive Science*, *42*(8), 2855–2889.

Portelas, R., Colas, C., Hofmann, K., & Oudeyer, P.-Y. (2020). Teacher algorithms for curriculum learning of deep rl in continuously parameterized environments. In *Conference on Robot Learning* (pp. 835–853). PMLR.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, *83*(2 PART 1), 304–308. https://doi.org/10.1037/h0028558

Powell, W. W., & Snellman, K. (2004). The knowledge economy. *Annu. Rev. Sociol.*, *30*, 199–220.

Prenzel, M., & Mandl, H. (1993). Transfer of learning from a constructivist perspective. In *Designing environments for constructive learning* (pp. 315–329). Springer.

Price, A. L. (2009). Distinguishing the contributions of implicit and explicit processes to performance of the weather prediction task. *Memory & Cognition*, *37*(2), 210–222.

Qin, W., Hu, Z., Liu, X., Fu, W., He, J., & Hong, R. (2020). The Balanced Loss Curriculum Learning. *IEEE Access*, *8*, 25990–26001.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, *1*(8), 9.

Rahmandad, H., Repenning, N., & Sterman, J. (2009). Effects of feedback delay on learning. *System Dynamics Review*, *25*(4), 309–338.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology:*

General, *118*(3), 219.

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*(1), 1–41. https://doi.org/10.1016/j.cogpsych.2004.11.001

Rescorla, R. A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Current Research and Theory*, 64–99.

Rich, A. S., & Gureckis, T. M. (2018). The limits of learning: Exploration, generalization, and the development of learning traps. *Journal of Experimental Psychology: General*, *147*(11), 1553.

Roads, B. D., Xu, B., Robinson, J. K., & Tanaka, J. W. (2018). The easy-to-hard training advantage with real-world medical images. *Cognitive Research: Principles and Implications*, *3*(1). https://doi.org/10.1186/s41235-018-0131-6

Rohde, D. L. T., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, *72*(1), 67–109.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, *35*(6), 481–498.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Rosenbaum, C., Klinger, T., & Riemer, M. (2017). Routing networks: Adaptive selection of non-linear functions for multi-task learning. *ArXiv Preprint ArXiv:1711.01239*.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386.

Rosenthal, J. S. (1995). Active learning strategies in advanced mathematics classes. *Studies in Higher Education*, *20*(2), 223–228.

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–349.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., … Bernstein, M. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*(3), 211–252.

Sadaghiani, S., Poline, J.-B., Kleinschmidt, A., & D'Esposito, M. (2015). Ongoing dynamics in large-scale functional connectivity predict perception. *Proceedings of the National Academy of*

Sciences, *112*(27), 8463–8468.

Saks, A. M. (2002). So what is a good transfer of training estimate? A reply to Fitzpatrick. *The Industrial-Organizational Psychologist*, *39*(3), 29–30.

Saville, C. W. N., Feige, B., Kluckert, C., Bender, S., Biscaldi, M., Berger, A., … Klein, C. (2015). Increased reaction time variability in attention- deficit hyperactivity disorder as a response-related phenomenon: evidence from single- trial event- related potentials. *Journal of Child Psychology and Psychiatry*, *56*(7), 801–813.

Schaul, T., Quan, J., Antonoglou, I., & Silver, D. (2015). Prioritized experience replay. *ArXiv Preprint ArXiv:1511.05952*.

Schuck, N. W., Gaschler, R., Wenke, D., Heinzle, J., Frensch, P. A., Haynes, J.-D., & Reverberi, C. (2015). Medial prefrontal cortex predicts internally driven strategy shifts. *Neuron*, *86*(1), 331–340.

Schulz, L. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, *16*(7), 382–389.

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology*, *43*(4), 1045.

Shah, B., Pattanayak, R. D., & Sagar, R. (2014). The study of patient henry Molaison and what it taught us over past 50 years: Contributions to neuroscience. *Journal of Mental Health and Human Behaviour*, *19*(2), 91.

Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., & Dean, J. (2017). Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *ArXiv Preprint ArXiv:1701.06538*.

Shrivastava, A., Gupta, A., & Girshick, R. (2016). Training region-based object detectors with online hard example mining. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 761–769).

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(4), 907.

Simon, H. A. (1957). Models of man; social and rational.

Şimşek, Ö., & Barto, A. G. (2006). An intrinsic reward mechanism for efficient exploration. In *Proceedings of the 23rd international conference on Machine learning* (pp. 833–840).

Smith, J. D., & Minda, J. P. (2000). Thirty Categorization Results in Search of a Model. *Journal of Experimental Psychology: Learning Memory and Cognition*, *26*(1), 3–27. https://doi.org/10.1037/0278-7393.26.1.3

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*(4), 342–353.

Soto, F. A., Gershman, S. J., & Niv, Y. (2014). Explaining compound generalization in associative and causal learning through rational principles of dimensional generalization. *Psychological Review*, *121*(3), 526.

Spiering, B. J., & Ashby, F. . (2008). *Initial training with difficult items facilitates information integration, but not rule-based category learning: Research article*. *Psychological Science* (Vol. 19). https://doi.org/10.1111/j.1467-9280.2008.02219.x

Stadler, M. A., & Frensch, P. A. (1998). *Handbook of implicit learning.* Sage Publications, Inc.

Stephan, K., Penny, W., Daunizeau, J., & Moran, R. (2009). Bayesian model selection for group studies. *Neuroimage*. Retrieved from https://www.sciencedirect.com/science/article/pii/S1053811909002638

Strange, W., Keeney, T., Kessel, F. S., & Jenkins, J. J. (1970). Abstraction over time of prototypes from distortions of random dot patterns: A replication. *Journal of Experimental Psychology*, *83*(3 PART 1), 508–510. https://doi.org/10.1037/h0028846

Sukhbaatar, S., Lin, Z., Kostrikov, I., Synnaeve, G., Szlam, A., & Fergus, R. (2017). Intrinsic motivation and automatic curricula via asymmetric self-play. *ArXiv Preprint ArXiv:1703.05407*.

Sun, R. (2008). *The Cambridge handbook of computational psychology*. Cambridge University Press.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, *24*(6), 837–848.

Terrace, H. S. (1964). Wavelength generalization after discrimination learning with and without errors. *Science*, *144*(3614), 78–80. https://doi.org/10.1126/science.144.3614.78

Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System*, *21*(3), 371–380.

Tulving, E. (1972). 12. Episodic and Semantic Memory. *Organization of Memory/Eds E. Tulving, W.*

Donaldson, NY: Academic Press, 381–403.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352.

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*, *42*(1), 1636–1643. https://doi.org/10.1111/ejn.12936

Van Vliet, P. M., & Wulf, G. (2006). Extrinsic feedback for motor learning after stroke: what is the evidence? *Disability and Rehabilitation*, *28*(13–14), 831–840.

Vance, R. J., & Colella, A. (1990). Effects of two types of feedback on goal acceptance and personal goals. *Journal of Applied Psychology*, *75*(1), 68.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. *ArXiv Preprint ArXiv:1706.03762*.

Vilidaite, G., Yu, M., & Baker, D. H. (2017). Internal noise estimates correlate with autistic traits. *Autism Research*, *10*(8), 1384–1391.

Vlach, H. (2014). The spacing effect in children's generalization of knowledge: allowing children time to forget promotes their ability to learn. *Child Development Perspectives*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/cdep.12079/full

Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*(2), 228.

Vygotskiĭ, L. S. (1978). *Mind in society: the development of higher psychological processes*. Retrieved from https://books.google.com/books?hl=en&lr=&id=Irq913lEZ1QC&oi=fnd&pg=PR13&dq=Mind+in+society:+the+development+of+higher+psychological+&ots=HbCqE3yhni&sig=LU2iePaLsmCMAVh84VzwzgCy_h8

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition*, *39*(5), 750–763.

Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2018). GLUE: A multi-task benchmark and analysis platform for natural language understanding. *ArXiv Preprint ArXiv:1804.07461*.

Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*(3–4), 279–292.

Weinshall, D., & Amir, D. (2020). Theory of curriculum learning, with convex loss functions. *Journal of Machine Learning Research*, *21*. Retrieved from http://arxiv.org/abs/1812.03472

Werbos, P. (1974). Beyond regression:" new tools for prediction and analysis in the behavioral sciences. *Ph. D. Dissertation, Harvard University*.

Whittington, J. C. R., & Bogacz, R. (2019, March 1). Theories of Error Back-Propagation in the Brain. *Trends in Cognitive Sciences*. Elsevier Ltd. https://doi.org/10.1016/j.tics.2018.12.005

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, *8*(3–4), 229–256.

Willshaw, D. (1994). Non-symbolic approaches to artificial intelligence and the mind. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, *349*(1689), 87–102.

Wilson, B. A., Baddeley, A., Evans, J., & Shiel, A. (1994). Errorless learning in the rehabilitation of memory impaired people. *Neuropsychological Rehabilitation*, *4*(3), 307–326.

Wilson, R. C., Shenhav, A., Straccia, M., & Cohen, J. D. (2018). The Eighty Five Percent Rule for Optimal Learning. *BioRxiv*, 255182. https://doi.org/10.1101/255182

Wozniak, P. A., & Gorzelanczyk, E. J. (1994). Optimization of repetition spacing in the practice of learning. *Acta Neurobiologiae Experimentalis*, *54*, 59.

Xu, Z., van Hasselt, H., & Silver, D. (2018). Meta-gradient reinforcement learning. *ArXiv Preprint ArXiv:1805.09801*.

Yan, F. F., Hou, F., Lu, H., Yang, J., Chen, L., Wu, Y., … Huang, C. B. (2020). Aging affects gain and internal noise in the visual system. *Scientific Reports*, *10*(1), 1–10. https://doi.org/10.1038/s41598-020-63053-0

Yang, G. J., Murray, J. D., Repovs, G., Cole, M. W., Savic, A., Glasser, M. F., … Pearlson, G. D. (2014). Altered global brain signal in schizophrenia. *Proceedings of the National Academy of Sciences*, *111*(20), 7438–7443.

Yang, T., & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature*, *447*(7148), 1075–1080.

Zaremba, W., & Sutskever, I. (2014). Learning to Execute. *ArXiv Preprint ArXiv:1410.4615*. Retrieved from http://arxiv.org/abs/1410.4615

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition*, *34*(2), 387–398.

Zhang, Cha, & Zhang, Z. (2014). Improving multiview face detection with multi-task deep convolutional neural networks. In *IEEE Winter Conference on Applications of Computer Vision* (pp. 1036–1041). IEEE.

Zhang, Chiyuan, Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, *64*(3), 107–115.

Zheng, Z., Oh, J., Hessel, M., Xu, Z., Kroiss, M., Van Hasselt, H., … Singh, S. (2020). What Can Learned Intrinsic Rewards Capture? In *International Conference on Machine Learning* (pp. 11436–11446). PMLR.

Zheng, Z., Oh, J., & Singh, S. (2018). On learning intrinsic rewards for policy gradient methods. *ArXiv Preprint ArXiv:1804.06459*.