A COMMON PROBABILISTIC FRAMEWORK EXPLAINING LEARNING AND GENERALIZATION IN PERCEPTUAL AND STATISTICAL LEARNING

Gábor Lengyel

Central European University
Department of Cognitive Science

In partial fulfilment of the requirements for the degree of Doctor of Philosophy in Cognitive Science

Supervisor: József Fiser

Secondary supervisor: Máté Lengyel

Budapest, August 2021

Declaration of Authorship

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or which have been accepted for the award of any other degree or diploma at Central European University or any other educational institution, except where due acknowledgement is made in the form of bibliographical reference.

The present thesis includes studies that appeared in the following articles:

1. Most parts of Chapter 1 was published in:

   Fiser, J. & Lengyel, G. (2019). A common probabilistic framework for perceptual and statistical learning [Computational Neuroscience]. *Current Opinion in Neurobiology*, *58*, 218–228. https://doi.org/https://doi.org/10.1016/j.conb.2019.09.007

   Author contributions: J.Fiser and G.Lengyel did the conceptualization. G.Lengyel performed the formal analysis and visualization. J.Fiser wrote the paper, with comments from G.Lengyel.

2. The study in Chapter 2 was published in:

   Lengyel, G. & Fiser, J. (2019). The relationship between initial threshold, learning, and generalization in perceptual learning. *Journal of Vision*, *19*(4), 28–28. https://doi.org/10.1167/19.4.28

   Author contributions: G.Lengyel and J.Fiser designed the study. G.Lengyel performed the experimental studies. G.Lengyel analyzed the experimental data. G.Lengyel and J.Fiser interpreted the results and wrote the paper.

3. The studies in Chapter 3 were published in:

Lengyel, G., Nagy, M. & Fiser, J. (2021). Statistically defined visual chunks engage object-based attention. *Nature Communications*, *12*(1), 272. https://doi.org/10.1038/s41467-020-20589-z

Author contributions: G.Lengyel, M.Nagy, and J.Fiser designed the study. G.Lengyel and M.Nagy performed the experimental studies. G.Lengyel analyzed the experimental data. All authors interpreted the results. G.Lengyel and J.Fiser wrote the paper, with comments from M.Nagy.

Lengyel, G., Žalalytė, G., Pantelides, A., Ingram, J. N., Fiser, J., Lengyel, M. & Wolpert, D. M. (2019). Unimodal statistical learning produces multimodal object-like representations (J. Diedrichsen, M. J. Frank, M. Landy & S. J. Gershman, Eds.). *eLife*, *8*, e43942. https://doi.org/10.7554/eLife.43942

Author contributions: G. Lengyel, D. M. Wolpert, M. Lengyel, J. Fiser and J.N. Ingram was involved in the conceptualization. G. Ẑalalytė, A. Pantelides and J.N. Ingram performed the experimental studies. D. M. Wolpert, M. Lengyel, G. Lengyel performed the formal analysis. D. M. Wolpert, M. Lengyel, G. Lengyel, G. Ẑalalytė and A. Pantelides analyzed the experimental data. G. Lengyel, D. M. Wolpert, M. Lengyel, J. Fiser and J.N. Ingram interpreted the results. D. M. Wolpert, M. Lengyel, J. Fiser and G. Lengyel wrote the paper, with comments from G. Ẑalalytė, A. Pantelides and J.N. Ingram.

Gabor Lengyel

Abstract

Sensory learning, the process of refining perception to improve interactions with the environment in a lasting manner, is traditionally divided into two learning types: perceptual and statistical learning. The two forms of learning have been treated separately in the literature in terms of paradigms, computational models, and neural correlates. However, recent experiments eliminated the strict distinctions between PL and SL paradigms by using more complex stimuli and tasks and as a result, they found overlapping computational and neural mechanisms between the two learning types. In the present thesis, I propose a common probabilistic framework that unifies the two forms of learning and can parsimoniously explain both classical findings seemingly supporting the separation between PL and SL and more recent results demonstrating strong interactions between the two forms of learning. I argue that Hierarchical Bayesian Modeling (HBM) that inherently combines sensory bottom-up and experience-based top-down processes in a normative manner can provide a suitable unifying framework for PL and SL. In particular, HBM provides higher flexibility for efficient generalization in situations with more complex, naturalistic tasks and stimuli, displaying a hallmark of human learning.

Inspired by the HBM framework, I conducted three empirical studies investigating learning and generalization in PL and SL paradigms and, after developing a computational model, I performed a simulation-based study exploring the interaction between PL and SL. In the first study, I investigated the relationship between initial performance, the amount of learning, and the extent of generalization in classical PL paradigms. I showed that (1) the previously found Weber-like relationship between initial performance and learning only shows properties of perception and does not reflect any characteristics of PL and (2) the extent of generalization was

proportional to the amount of learning. Studies 2 & 3 targeted SL focusing on how learning regularities in the stimuli influences perceptual organization. By combining the classical SL with the classical object-based attention paradigms, I showed that statistically defined chunks learnt during SL elicit object-based perceptual & attentional effects similar to what real objects do. Next, using visual and haptic stimulation in two SL experiments, we found that participants generalized the statistically defined chunks learnt in one modality to the other modality without any learning in the other modality. These findings suggest that, relying on statistical properties, participants automatically build abstract and amodal representations of chunks that influence the segmentation of the sensory input into perceptual units. Finally, I used computational modelling to study the interaction between PL & SL in roving paradigms and developed a unifying Bayesian Statistical Perceptual Learning model that can capture behavior in both classical and roving PL experiments. In the model, the context of the trials are inferred and the temporal transition model between the contexts is gradually learned via SL, which in turn supports the PL process modelled as an efficient resource allocation for encoding the stimuli. This interaction between PL & SL in the model could replicate the wide range of results found in roving paradigms.

Together, these results pave the road to a novel understanding of learning in vision and the concept of perceptual "objects".

Acknowledgements

I am thankful to the following people for providing help, support, and guidance in my research.

First, I am grateful to my primary supervisor, József Fiser, who helped me designing my experiments, interpreting the results, and writing scientific papers. He also supported all of my applications to summer schools and research visits which enabled me to gain research experience in several prestigious labs outside of CEU.

Second, I would like to thank my secondary supervisor, Máté Lengyel, for technical advice on computational modelling in all of my research projects, for helping in designing, analyzing, and writing the third empirical study of this thesis (see Section 3.3), and for supporting my applications to summer schools and research visits.

Third, I owe special thanks to Ádám Koblinger for providing useful advice on computational modelling and data analysis, and to Márton Nagy for helping in designing and conducting experiments and for his direct contribution to the second study of this thesis (see Section 3.2).

I am also thank full to all of my co-authors who contributed in the third empirical study of this thesis (see Section 3.3), especially to Daniel M. Wolpert for designing, analyzing, and writing the study, to Goda Žalalytė and Alexandros Pantelides for conducting and analyzing the experiments, and to James N. Ingram for helping in designing, and coding the experiments, and interpreting the results.

I am also grateful to all former and current members of our lab, and to the members of the Department of Cognitive Science for the useful scientific discussions and providing a supportive work environment.

I would like to thank Réka Finta for providing rapid and smooth help during my PhD program.

I am also thankful to the CEU, to the Cortical Circuit Dynamics Group at IDIBABPS in Barcelona, to the University of Cambridge, to the Facebook Reality Labs, and to CoSMo (University of Minnesota), CCN (NYU Shanghai), and CBMM (Marine Biological Laboratory) Summer schools.

Finally, I would like to thank my family, my friends for their constant support and most of all to my grandmother who encouraged me to pursue a career in academia.

Contents

List of Figures

Chapter 1

A Unified Framework for Sensory Learning

In everyday life, learning refers to the intentional process of acquiring new knowledge, understanding and motor skills either alone or with the help of an instructor. When we think about learning, it typically involves the use of language and self-initiated movements, e.g. when someone is studying chess openings from a book or learning to do a hand stand by carefully executing the instructions of a personal trainer. However, most learning processes in our brain do not involve language or intentional motor movements, but rely entirely on unconscious, sensory-motor processes. Although we are not aware of it, our brain engages in sensory and motor learning all the time as we unconsciously adapt to changes in our environment, handle new objects, and refine our perception and motor skills in old environments. In all of these cases, the brain updates its representation using the information that the brain receives from the senses.

The most important feature of any learning is that it causes relatively permanent changes in brain processes. This characteristic distinguishes learning from sensitization (i.e. habituation

or adaptation) and priming, which represent short-term enhancing (priming) and hindering (adaptation) modulations in brain processes due to recent stimulation. Another fundamental aspect of learning is the extent to which it generalizes to new situations. Acquired knowledge usable later only for one specific input pattern or context is not useful in a constantly changing environment since the exact same input never occurs twice under natural conditions. For example, identifying the same object under different viewing conditions requires generalizing the learned appearance of the object to other retinal inputs, viewing angles, and contrast as well (DiCarlo et al., 2012; Serre et al., 2007). Indeed, human learning allows powerful generalization. From language acquisition (Pinker, 1998; Xu & Tenenbaum, 2007) through perception (Carey, 2009a; Gopnik et al., 2004) to motor learning (Braun et al., 2004; Braun et al., 2010), we develop abstract representations through inductive learning that can be applied to a much wider range of sensory events than the sensory data which it is based on (Lake et al., 2015).

In this thesis, I will investigate generalization effects in sensory learning. Sensory learning is the process of adapting or refining perception to our environment in a lasting manner. This form of learning does not involve motor learning, nor any language related, symbolic knowledge acquisition. Initially, researchers have focused on investigating human sensory learning in early childhood (Gibson, 1967). Inspired by the results of these developmental studies (Gibson, 1967), subsequent research systematically documented the various forms and characteristics of sensory learning in a wide range of tasks in adults (Ball & Sekuler, 1982; DeValois, 1977; Fendick & Westheimer, 1980; Fiorentini & Berardi, 1980; Ramachandran & Braddick, 1973; Vogels & Orban, 1985). Based on those studies, it became clear that after extensive practice, even the most basic, low-level aspects of perception such as contrast perception can be permanently changed. These earlier studies also established that, similar to motor learning, sensory

learning is most dominant in early childhood, when one learns to perceive the environment. Nevertheless, this learning continues throughout the entire lifespan although to a lesser degree than in infancy (Fahle & Poggio, 2002).

From these early investigations focusing on the issue of how perception changes relatively permanently after extensive practice in adulthood, an entire new field emerged which is named Perceptual Learning (PL). PL traditionally investigates low-level, simple, perceptual, several-day-long learning tasks with rigorous feedback (Fahle & Poggio, 2002; Sagi, 1994). The most dominant characteristic of PL is specificity; the lack of generalization of learning to other stimulus features and tasks (Fahle et al., 2002). This specificity of learning in PL tasks had attracted great interest among researchers since almost all other forms of human learning showed at least some degree of generalization.

A decade later, another type of learning was introduced to the domain of sensory learning, which originated from classical conditioning and associative learning, and which was called Statistical Learning (SL). SL explores how spatial and temporal patterns (i.e. correlations among elements) in the sensory input are learnt without any explicit feedback revealing the patterns (Aslin, 2017; Saffran & Kirkham, 2018). SL goes beyond classical conditioning and associative learning by (1) studying more complex structures in the stimuli and not just co-associations, and (2) investigating the representation and the computational mechanisms of learning the statistical structure in the stimuli. In contrast to PL, the learnt structure during SL can be applied to new stimuli and tasks (Fiser & Aslin, 2005; Marcus et al., 1999) demonstrating powerful generalization effects; the marker of human learning.

The literature investigating PL and SL in the domain of sensory learning traditionally has been keeping a strict separation between these two learning types. This separation of the two forms of learning was reasonable given the seemingly large differences in the traditional PL

and SL paradigms, characteristics, and results (see sections 1.1 and 1.2 for details). However, in more recent studies, the methods and results of the two domains of learning have started to converge and researchers were able to demonstrate a wide range of generalization effects in both in PL and SL pointing toward somewhat overlapping learning mechanisms (see section 1.3 for details). Therefore, maintaining the original strict separation between the two learning types is not parsimonious any more.

In the present thesis, I will investigate the interaction between PL and SL in adults focusing on the forms of generalization in the two types of learning. I will argue that sensory learning needs a new framework that treats PL and SL uniformly and jointly in order to seamlessly integrate recent empirical findings showing a large overlap between the neural substrates and the computational principles of PL and SL. In the introduction, first, I will review the two types of learning summarizing the most important behavioural findings, computational models, and neural correlates of PL and SL. Second, I will present recent studies that suggest a vanishing distinction between PL and SL both in terms of methods and results. Third, I will provide a new framework that offers a parsimonious account for previous findings and suggests new paradigms and computational models for investigating PL and SL together. Finally, I will provide a brief outline of the thesis summarizing the structure and the topics in the following four chapters. A shorter, more condensed version of the parts of this first chapter was published in Fiser and Lengyel (2019). Note that the present chapter focuses on reviewing studies in the visual domain, and there exists a substantial body of studies in sensory learning targeting other domains that are not covered in this chapter.

1.1 Perceptual Learning (PL)

Perceptual learning (PL) is defined as improvement in simple sensory tasks with extensive practice (Fahle & Poggio, 2002; Sagi, 1994). In a typical PL task participants see two artificially generated grating stimuli (e.g. two Gabor patches) sequentially and they have to decide whether the first or the second stimulus had a higher value with respect to a low-level perceptual feature such as contrast, orientation, or brightness (Fig. 1.1A). This paradigm is called the two alternative forced choice (2-AFC) discrimination task and a typical finding is that with practice, participants learn to see and able to discriminate smaller and smaller differences between the two stimuli' feature values, e.g. contrast levels (Fig. 1.1B). PL can be demonstrated in real life tasks too, e.g. when medical students in radiology learn to detect bad tissue and to discriminate from each other different types of soft tissue in medical images or when, after long extensive practice, an ornithologist, is able to tell apart two bird species (Fig. 1.1C).

1.1.1 Typical behavioral results

PL was demonstrated using artificial stimuli, among other tasks, in contrast (Adini et al., 2004; Yu et al., 2004) and motion detection (Ball & Sekuler, 1987), orientation (Fiorentini & Berardi, 1980) and texture discrimination (Ahissar & Hochstein, 1997; Karni & Sagi, 1991), hyperacuity (Spang et al., 2010) and stereoscopic vision (O'Toole & Kersten, 1992). PL was also demonstrated in tasks using naturalistic stimuli in a number of expertise-learning studies (Devillez et al., 2018; Tanaka et al., 2005). Extensive practice typically amounts to 5-14 days of repetitive exposure over 1-2 hours (Jeter et al., 2010). Sleeping across the days is necessary for PL since it significantly alters the amount of learning due to consolidation (Karni et al., 1994; Miyamoto et al., 2016), and the changes remain in effect for days, months, even years (Karni & Sagi, 1993). While in a few studies, feedback on the correctness of the observer's response

during trials was not provided (Watanabe & Sasaki, 2015), typically, there is feedback, and it is crucial for improving (Aberg & Herzog, 2012) or even permitting learning (Shiu & Pashler, 1992).

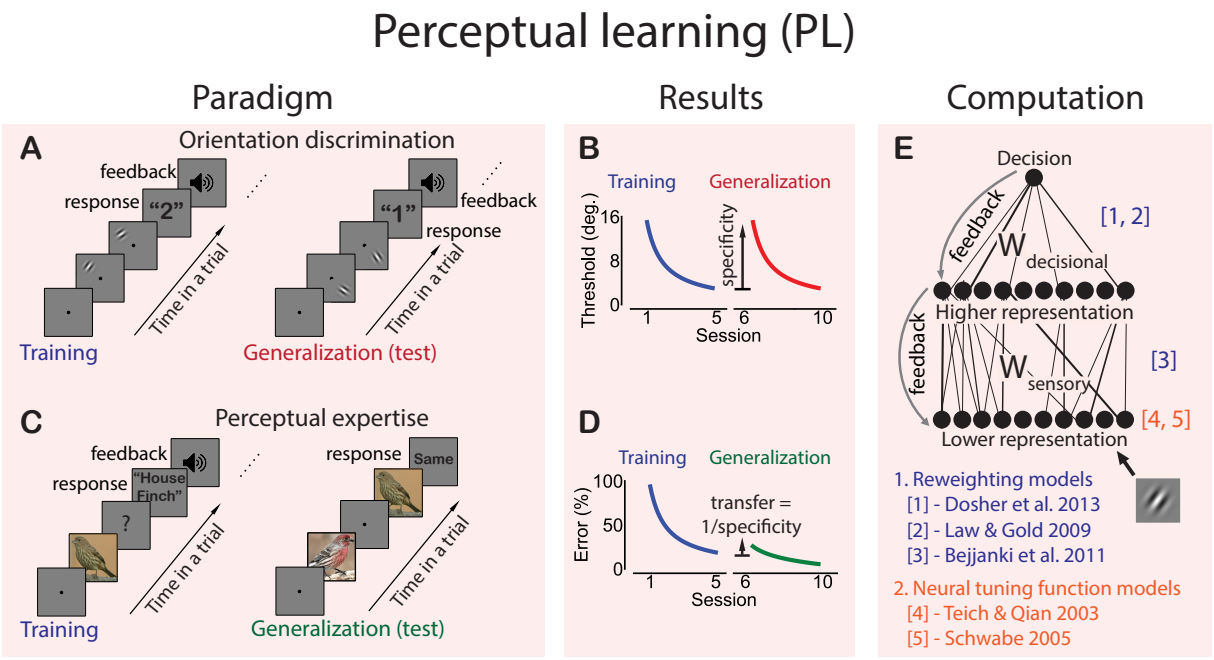


Figure 1.1: Classical perceptual learning. The paradigms (**A & C**), typical behavioral results (**B & D**), and computational frameworks (**E**) of perceptual learning. **A, B:** Classical orientation discrimination task with the corresponding performance improvement in the trained condition (drop in blue curve) and specificity (i.e. a lack of transfer of performance to a different condition, initial jump in red curve). **C, D:** Perceptual expertise task of bird species discrimination showing both improving performance with trained birds (blue curve) and generalization to previously unseen birds (transfer i.e. no initial jump in green curve). **E:** Structure and references of the dominant computational models in PL assuming tuning changes in the representational units (orange) or re-weighting of representation-to-decision connections (blue). Adapted with permission from J. Fiser and G. Lengyel.

The amount of learning is usually measured in improvements of a threshold indicating a change in perceptual sensitivity (Fahle et al., 2002). The threshold in detection tasks represents the absolute value of the intensity with which a perceptual feature needs to be presented so that it can reliably be detected in background noise. For example, detection threshold would be the minimum level of contrast of an image about a dog at which the dog can be detected in the image . In the case of a discrimination task, threshold represents the smallest change in the intensity of a feature that can be detected. For example, the smallest angular difference between

two lines that can reliably be reported would be a discrimination threshold in an orientation discrimination task.

There are several hallmarks of perceptual learning that cast this type of learning as a low-level phenomenon. The first is the specificity of learning: the acquired improvement in performance does not hold when conditions are altered (Fig. 1.1B). This is the opposite of generalization. Examples of such alterations are the stimulus being presented at a different location (Fahle & Morgan, 1996; Schoups et al., 1995), orientation (Crist et al., 1997), spatial frequency (Fiorentini & Berardi, 1980), paired with different background (Crist et al., 1997) or seen through a different eye (Schoups et al., 1995). Imagine, one has learnt to discriminate orientations from a reference orientation in the task shown in Fig. 1.1A. If this observer sees a new reference orientation or if the stimuli are placed to a different location on the screen, the observer has to relearn the entire task regardless of the fact that the previous learning on the orientation discrimination task has been accomplished. Because of the specificity of PL, researchers argued that the learning takes place in the early visual cortex having small receptive fields that are more selective to certain retinal locations and features. Eye-specificity has been used heavily to argue for a low-level origin of PL: since merging of monocular representations happens in V1, eye-specific differences require learning also to occur in the primary visual cortex (Schoups et al., 2001).

1.1.2 Neural correlates

In line with the argument of PL being a low-level learning process, investigations of possible substrates of PL focused originally on early areas of the perceptual hierarchy and on firing responses of specific individual neurons (Fig. 1.2 in red). Indeed, single cell recording studies showed an increased slope of the orientation tuning function at the trained orientation in awake

monkeys' primary visual cortex (V1) (Schoups et al., 2001) and an increased contrast sensitivity of trained spatial frequencies in anesthetized cats' V1 after PL (Hua et al., 2010). More recently, human functional magnetic resonance imaging (fMRI) studies also found training-specific increase of orientation selectivity in human V1 blood-oxygen-level dependent (BOLD) signal (Jehee et al., 2012) and increased BOLD signal in the human lateral geniculate nucleus (LGN) in response to low-contrast images (Yu et al., 2016).

However, a number of studies reported a more widespread neural effect of learning both in terms of cortical location and by the type of change in neural behavior. PL caused reduced variability of single cell responses to oriented gratings in V4[1] and medial superior temporal area (MST) of awake monkeys (Adab & Vogels, 2011; Gu et al., 2011) as well as in fMRI BOLD signals in posterior inferotemporal area (PIT) (Adab et al., 2014). Even in V1, PL resulted in changes not only in the tuning characteristics of individual cells, but in their variability, gain modulation, their population correlational structure or their functional connectivity pattern (LeMessurier & Feldman, 2018). The combination of such changes were proposed to serve a combined signal enhancement and background suppression to support both encoding and readout of sensory information (Yan et al., 2014). In agreement with this, PL induced improvement not only in single cell encoding (Adab & Vogels, 2011), but by reducing noise correlations, and enhancing readout in area V4 of behaving monkeys during a 2-AFC contrast discrimination task (Ni et al., 2018; Sanayei et al., 2018). Imaging studies also found that the effects of PL extended far beyond simple mean changes of neural response in a dedicated cell population. Using dynamic motion stimuli, several fMRI and transcranial magnetic stimulation (TMS) studies found that the functional specializations of cortical areas could change due to PL: depending on the global nature of the actual stimulus type, the causal link between activ-

[1] area in the extrastriate visual cortex; located anterior to secondary visual area [V2] and posterior to posterior inferotemporal area [PIT]

ity of the middle temporal visual area (MT) and motion processing could increase or decrease (Chen et al., 2016; Liu & Pack, 2017). At the higher level of face discrimination, PL positively correlated with the amount of reduction in variability measured in the neural patterns of multi-voxel pattern analysis (MVPA) in the left fusiform cortex (Bi et al., 2014). Similarly, a classification image method applied to fMRI data in the higher human lateral occipital cortical area (LO) showed the emergence of a refined template, which combined signals from the most informative parts of the input (Kuai et al., 2013). These findings point toward a complex influence of PL on neural coding attributed customarily to various top-down effects (Gilbert et al., 2001; Vogels, 2010).

1.1.3 Computational models

Classical computational studies of PL used the feed-forward/top-down dichotomy and biologically plausible neural network approaches to model PL (Abbott & Regehr, 2004; Tsodyks & Gilbert, 2004) (Fig. 1.1E). Such models applied to orientation discrimination in PL established that feedforward methods are insufficient to capture human behavior, but recurrent networks with lateral and top-down connections could replicate some of the hallmark characteristics of PL (Schwabe, 2005; Teich & Qian, 2003). In the past two decades, however, a more abstract and more psychophysics-oriented modeling scheme called the "reweighting model" emerged as the dominant computational framework of PL (Dosher & Lu, 2017; Dosher et al., 2013) (Fig. 1.1E). According to this approach, changes due to learning do not occur in early retinotopic sensory representations, but rather in the weight structure from sensory representations to decision units. Various psychophysical and neurophysiological findings are compatible with this proposal, which also strengthens the link between PL and associative learning schemes with reinforcement (Law & Gold, 2009, 2010). Several papers expanded the basic reweighting

model by adding realistic learning rules to it (Petrov et al., 2005), linking it to reinforcement learning (Roelfsema et al., 2010; Watanabe & Sasaki, 2015), and to hierarchical neural networks (Dosher et al., 2013), where the rate of learning at various levels of the hierarchy could be controlled by confidence due to task difficulty (Talluri et al., 2015).

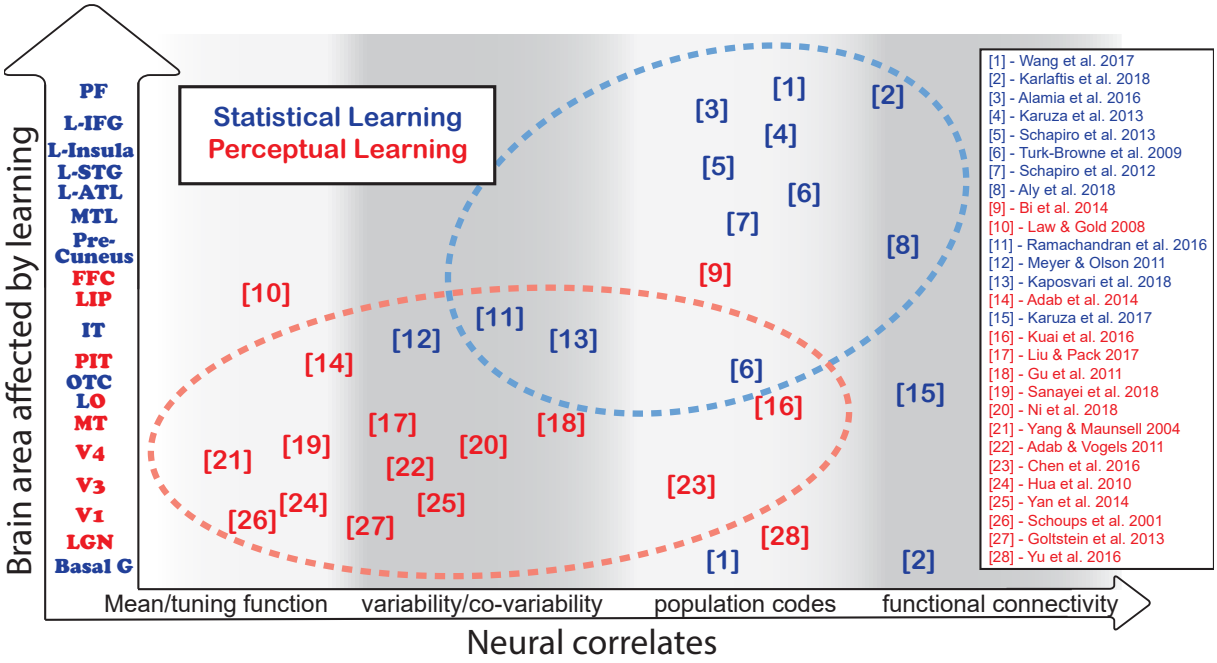


Figure 1.2: Neural correlates of perceptual and statistical learning. Reports on neural correlates of PL (red) and SL (blue) ordered along two relevant dimensions: the complexity of the reported neural correlate modulated by learning (x axis), and the rough position of the investigated brain area within the cortical hierarchy (y axis) colored in red/blue according to which learning was found to influence the area predominantly. Dashed areas indicate typical combinations of neural correlates and involved areas of PL (red) and SL (blue). PF: prefrontal cortex. L-IFG: left inferior frontal gyrus. L-Insula: left insula. L-STG: left superior temporal gyrus. L-ATL: left anterior temporal lobe. MTL: medial temporal lobe. Precuneus: portion of the superior parietal lobule on the medial surface. FFC: fusiform face complex. LIP: lateral intraparietal cortex. IT: inferior temporal cortex. PIT: posterior inferotemporal area. OTC: occipitotemporal cortex. LO: lateral occipital cortical area. MT: middle temporal visual area. V4: area in the extrastriate visual cortex. V3: third visual complex. V1: primary visual cortex. LGN: lateral geniculate nucleus. Basal G: basal ganglia. Adapted with permission from J. Fiser and G. Lengyel.

1.2 Statistical Learning (SL)

Statistical learning (SL) refers to the type of representational learning that is purely observational without any task or feedback, which automatically and implicitly re-represents repeatedly

appearing spatial and temporal patterns in the sensory input (Aslin, 2017; Saffran & Kirkham, 2018) (Fig. 1.3). Originally introduced in the domain of language learning for solving the problem of word segmentation (Saffran et al., 1996), statistical learning has been later predominantly investigated in the domain of vision (Fiser & Aslin, 2001; Kirkham et al., 2002). A large body of statistical learning studies, focusing on the domain of language development (see Erickson and Thiessen (2015), Newport (2016) and Saffran and Kirkham (2018)) will not be discussed here because this thesis focuses on the domain of vision.

SL paradigms can be classified broadly into two categories: spatial and temporal SL. In a typical spatial SL paradigm participants are exposed, one by one, to a sequence of unique multi-element visual scenes. Unbeknownst to the participants the visual scenes are generated according to a rule creating a consistent statistical structure between the spatial allocations of the elements across the scenes (see the inventory defining the spatial structure between the shapes in Fig. 1.3A). During the spatial SL paradigm, participants first passively observe the visual scenes after each other. After this exposure phase, they complete a familiarity test that measures whether they learnt the spatial structure and created a representation about the regularly occurring spatial allocations of the elements. Specifically, in the familiarity test they are asked to judge which of two novel visual scenes is more familiar to them where the elements in one of the two novel scenes were generated in accordance with the statistical structure hidden in the exposure phase while the elements in the other scene were randomly generated (see Generalization (test) in Fig. 1.3A). Usually, participants find the novel scenes generated according to the structure observed during the exposure phase more familiar demonstrating that they learnt the statistical structure in the visual scenes during the exposure phase (Fiser & Aslin, 2001) (Fig. 1.3B). In contrast, in a typical temporal SL paradigm participants see a sequence of scenes with only one element (instead of multi-element scenes) during the exposure phase.

There is a hidden temporal statistical structure in the sequence of scenes defining how often an element follows another element (see the inventory defining the temporal structure between the shapes in Fig. 1.3C). In the same way as in the spatial SL task, after the passive exposure to the sequence of the scenes with one element, participants complete a familiarity test in which they judge whether a short sequence from the exposure phase or a short sequence with randomly ordered elements is more familiar to them (see Generalization (test) in Fig. 1.3C). Participants find the sequence generated according to the temporal structure observed during the exposure phase more familiar suggesting that they created a representation about the temporal structure (Turk-Browne et al., 2005) (Fig. 1.3D).

SL is rarely demonstrated in real life tasks. However, there are plenty of studies showing that our perceptual system is adapted to the statistical structures in our environment (Bertenthal, 2001; Olshausen & Field, 1996), and developmental studies suggests that a large part of this adaptation happens during early childhood (Bertenthal, 1996; Spelke, 1990). Studies using more realistic stimuli and tasks demonstrated some form of SL in video games (Green et al., 2010a; Green et al., 2010b) and in image/object categorization tasks (Austerweil & Griffiths, 2011).

1.2.1 Typical behavioral results

Initial results in SL established that adults and infants alike demonstrate spatial and temporal statistical learning based on both joint and conditional probabilities as well as higher-order embedded structures of previously unknown inputs (Bulf et al., 2011; Fiser & Aslin, 2002, 2005; Schapiro et al., 2013; Slone & Johnson, 2018) (Fig. 1.3A & C). In spatial SL, joint probability refers to the probability of certain elements co-occur together in a fixed spatial allocation relative to each other. In temporal SL joint probability means the probability of two

or more elements appearing in a particular consecutive order.

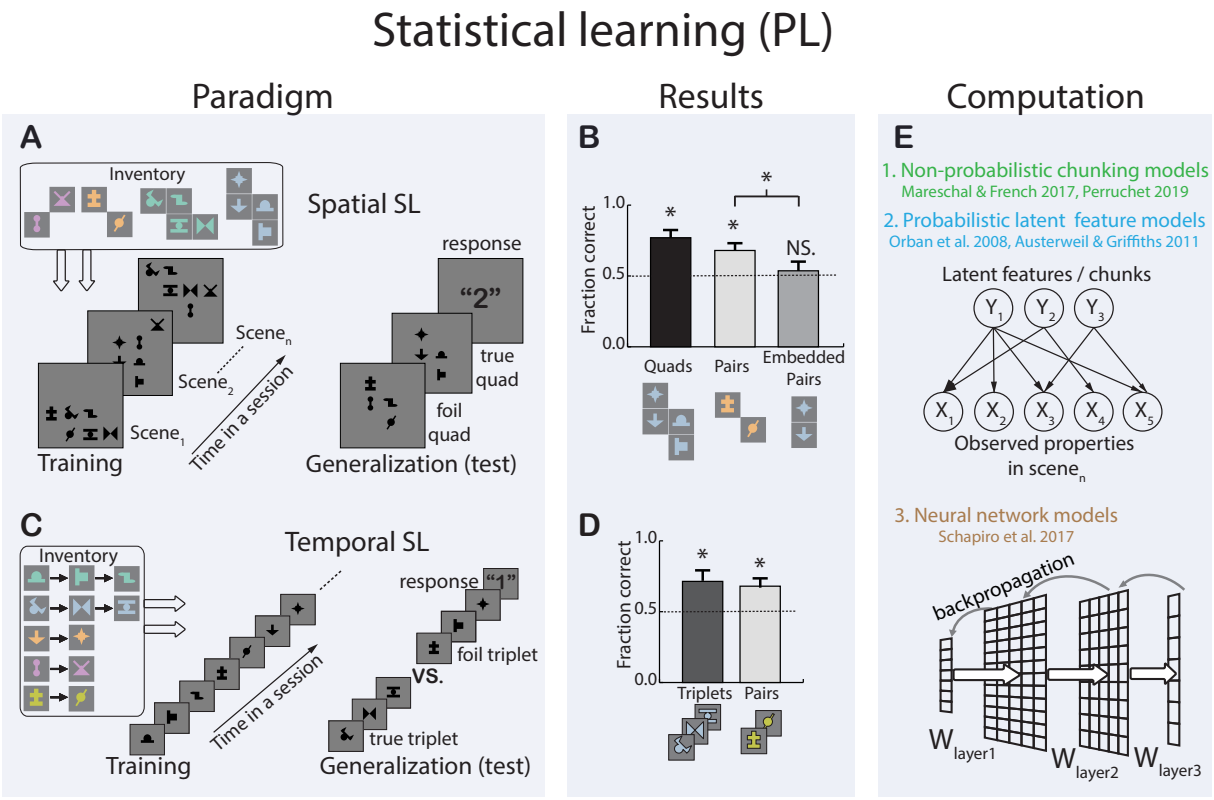


Figure 1.3: Classical statistical learning. The paradigms (**A & C**), typical behavioral results (**B & D**), and computational frameworks (**E**) of statistical learning. **A, B:** Classical spatial visual SL task with the inventory, the composed set of training scenes, the segmented substructures of the training scenes ("chunks") vs. random shape combinations used as test scenes, and the corresponding familiarity performance with the tests scenes indicating generalization of learning. **C, D:** Same as A & B but with classical temporal visual SL task using a long temporal chain of shape images from shape triplets as a training sequence and shape triplets presented consecutively as test stimuli in the familiarity test. **E:** Structure and references of the dominant computational models in SL based on non-probabilistic (green) and probabilistic (turquoise) latent chunk learning, and on biologically and computationally motivated connectionists learning (brown). Adapted with permission from J. Fiser and G. Lengyel.

In contrast, conditional probability quantifies the extent to which the appearance of an element can be predicted from the appearance of another element, thus it is often called predictability. In spatial SL, conditional probability is the occurrence probability of an element at a certain location given another element is present regardless of their joint co-occurrence frequency (see Fiser and Aslin (2001)). In temporal SL, conditional probability refers to the probability of an element following another element given that the other element was present, again regardless of how many times those two elements appeared in a particular consecutive order (see Fiser

and Aslin (2002)). Finally, higher-order embedded structures in spatial SL can be demonstrated with larger multi-element chunks containing smaller, embedded chunks as parts, e.g. two pairs of shapes (two smaller chunks) appearing separately as pairs, but also together as a quad (large chunk containing two smaller chunks as parts) frequently (see Fiser and Aslin, 2005). In a temporal SL paradigm, higher-order structures can be shown with generating a sequence of elements from a higher order Markov model, e.g. with community structure where the elements of the sequence are generated from a network with hubs (see Schapiro et al. (2013)).

The results demonstrating spatial and temporal statistical learning based on both joint and conditional probabilities, and higher-order embedded structures have been extended to various modalities (visual, auditory, tactile) (Conway & Christiansen, 2005; Glicksohn & Cohen, 2013; Lengyel et al., 2019; Ongchoco et al., 2016), to different stimulus complexities (Brady & Oliva, 2008; Turk-Browne et al., 2008), and to other animals species (Castro et al., 2018; Rosa-Salva et al., 2018; Santolin et al., 2016; Santolin & Saffran, 2018; Toro & Trobalón, 2005). This ubiquitousness fueled the proposal that statistical learning is a domain-general process that might serve as the fundamental learning method for acquiring internal representations of the environment (Fiser, 2009; Lengyel et al., 2019) even though some auxiliary domain-specific constraints might exist (Frost et al., 2015). SL is automatic and persists for days (Kim et al., 2009), sleep does not improve it (Nemeth et al., 2009; Simor et al., 2018) and while attention can influence SL (Turk-Browne et al., 2005), it is not required for successful learning (Musz et al., 2014). Statistical learning has also been linked to or contrasted with higher level abstract concept learning (Altmann, 2017) and rule learning (Marcus et al., 1999; Saffran et al., 2007). Rule learning differs from SL only by the level of abstraction at which the structure exists in the stimuli. For example, learning the rule of AAB means that two repetitions is followed by an alternation but A and B can represent any sensory event such as a syllable (Marcus et al., 1999)

or an image (Saffran et al., 2007).

1.2.2 Neural correlates

In contrast to PL, the neural manifestation of SL has been typically searched for at higher levels of cortical representation, often within the paradigm of temporal statistical learning, and the results were interpreted as sensitivity to predictability (Fig. 1.2). Regarding neural correlates, single cell firing responses in the inferotemporal (IT) cortex is known to be modulated by the transitions between images violating the sequence that monkeys were exposed previously (Kaposvari et al., 2018; Meyer & Olson, 2011). Similarly to human behavioral results, these IT cell responses also capture contingencies between images as reflected by the sensitivity of responses to the conditional probabilities between images (Ramachandran et al., 2016). At a larger scale, various methodologies and indicators were used to demonstrate widespread neural effects of SL including changes in electroencephalography (EEG) N400 signal (Abla et al., 2008) and entrainment (Batterink & Paller, 2017), fMRI activity in specific areas of the medial temporal lobe, lateral occipital cortex, hippocampus and frontal gyrus (Karuza et al., 2013; Schapiro et al., 2012; Turk-Browne et al., 2009), temporary or permanent disruptions of activity in higher cortical structures (Alamia et al., 2016; Schapiro et al., 2014), and anatomical measures of cortical thickness (Finn et al., 2018).

Beyond simply demonstrating effects of SL, a few studies focused on the more detailed link between the statistical structure of the input and its neural correlates during SL. Using the method of EEG mismatch negativity (MMN), the magnitude of the MMN signal was found to be inversely proportional with the auditory transitional probabilities (Koelsch et al., 2016). Using adaptation and multivoxel pattern analysis of human fMRI data and event sequences with higher-order cluster structure as stimuli for temporal SL, researchers found that the higher-order

chunk structure of the input was reflected in the neural activity of the anterior temporal lobe, insula and superior temporal gyrus (Schapiro et al., 2013). Moreover, the representation of natural scene categories in the anterior visual cortex is based on object co-occurrence statistics, thus both the presented scenes and the specific objects within can be decoded from the fMRI BOLD activity (Stansbury et al., 2013). Finally, both temporal and spatial SL studies using fMRI found that statistical structures defined on longer time-scales with regularities beyond immediate temporal transitions can be best detected in changes of the functional connectivity of neural responses (Aly et al., 2018; Karuza et al., 2017). In fact, diffusion tensor imaging (DTI) and fMRI results of sequential SL suggest that observers adopt different strategies, either learning exact sequences or selecting the most probable output. These two strategies involve the change in functional connectivity in two different sets of brain structures involving the caudate and the hippocampus for temporal versus the prefrontal, cingulate and basal ganglia for spatial SL (Karlaftis et al., 2018; Wang et al., 2017).

1.2.3 Computational models

Traditional computational studies of SL, especially of temporal SL, heavily mix general implicit associative learning, transitional probability (TP) counting methods and "chunk-learning" while using the classical connectionists framework (Mareschal & French, 2017; Perruchet, 2019) (Fig. 1.3E). The TP and associative learning models assume that only pair-wise correlations between the elements are learnt while chunking models assume that higher order structures, i.e. "chunk" are formed during SL. The emergent consensus is that TP-counting models are not sufficient to capture human learning even in temporal SL, but the different variants of chunk-learning can produce very similar results under simple input statistics (Perruchet, 2019). Only chunk-based learning models are able to capture the full variety of empirical findings in spatial

SL paradigms too (Orbán et al., 2008). Finally, the biologically more realistic neural network models focused on reconciling SL and rapid learning of episodic memories with known neurophysiological data (Schapiro et al., 2017).

1.3 Diminishing differences between PL and SL

While earlier studies have already found evidence indicating an overlap between the neural substrates and computational features of PL and SL (Dosher & Lu, 2017; LeMessurier & Feldman, 2018; Watanabe & Sasaki, 2015), more recent reports greatly accelerated this convergence due to the increasing similarity in stimulus complexity and task specificity between experiments conducted in the two domains (Fig. 1.4).

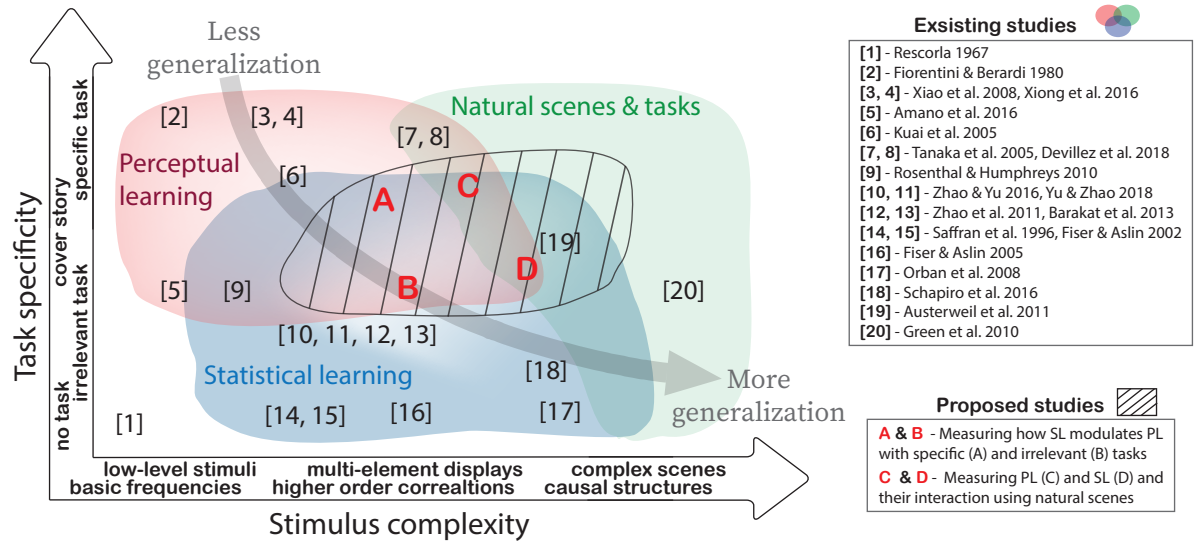


Figure 1.4: Vanishing differences between perceptual (PL) and statistical learning (SL). The relationships between PL (pink area) and SL (blue area) mapped onto the two dimensions of stimulus complexity (x axis) and task specificity (y axis). In recent studies [6,9,10-13,19] using more complex stimuli and a larger variability in the selected task that can create more natural conditions (green area), the classical separation between PL and SL waned. However, a systematic exploration on the integration of PL and SL (striped area) with specific new paradigms (A,B,C & D) still awaits. Bracketed numbers indicate references for previous studies, while letters indicate proposed new experiments (see legend on the right). Adapted with permission from J. Fiser and G. Lengyel.

In the domain of PL, it has been firmly established by now that PL induces changes not only

in V1 but in a large set of brain regions and it influences post-sensory processes as well (Diaz et al., 2017; Maniglia & Seitz, 2018). PL is task- and context-specific (Li et al., 2004), it appears to share common neural mechanisms with decision making processes in monkeys (Law & Gold, 2008, 2010) and humans (Kahnt et al., 2011), and both exogenous and endogenous spatial attention affect it (Donovan & Carrasco, 2018; Donovan et al., 2015). Even pure mental imagery without any sensory input can induce PL (Tartaglia et al., 2009c). Using a "double-training" learning paradigm, various studies reported enhanced or complete transfer of the learned ability to a new condition (Wang et al., 2014; Xiao et al., 2008) not only across different locations but across different physical properties that share "conceptual level" similarities (Wang et al., 2016). PL was enhanced when trials from multiple versions of the same task were delivered in a fixed order (Kuai et al., 2005). Transfer of learning depended on the precision of the transfer test, not only of the original training task (Jeter et al., 2009), and in general, the relationship between the type of the training and test tasks determined the success of generalization (Chang et al., 2013; Chang et al., 2014; Lengyel & Fiser, 2019). In addition, increasing stimulus complexity also facilitates generalization (Hussain et al., 2012). Higher-level generalization in PL has been investigated with training to play video games and learning was manifested not by simply having better attention, but by improved ability to generate templates for task learning (Green et al., 2015; Green et al., 2010a). Such structure-learning revealed by a faster learning rate could occur independently from the traditional immediate transfer in performance during PL (Kattner et al., 2017).

In the domain of SL, there has also been a steady progress of expanding and concretizing the areas and the extent to which SL influences or changes perceptual processes (Fig. 1.4). SL interferes with the process of extracting summary statistics of scenes (Zhao et al., 2011),

attention is spontaneously biased to structures identified implicitly by SL (Zhao et al., 2013; Zhao et al., 2011; Zhao & Yu, 2016), and SL reduces perceived numerosity (Zhao et al., 2011; Zhao & Yu, 2016). SL enhances memory for element of learned triplets and reduces memory for inserted distractors (Otsuka & Saiki, 2016), alters the internal representation of pair elements based on their predictability (Barakat et al., 2013), and it can create novel object associations based on transitive relations (Luo & Zhao, 2018). Importantly, these kinds of associations do not only establish novel links between the identity of elements, but also influence perception of features across elements. For example, after learning that two elements belong to the same pair, seeing one of them at a different size will influence the observer's perception of the size of the other element (Yu & Zhao, 2018). These effects have been typically conceptualized as top-down influences reaching down to even the most basic attributes, such as motion perception (Sotiropoulos et al., 2011) or rivalry (Piazza et al., 2018), and they can be manifested neurally at the lowest level of cortical representations (Köver et al., 2013) similarly to findings in PL.

The above summary suggests that, in contrast to their original conceptualization, PL and SL share characteristics in almost every domain. Both of them can influence various neural metrics at multiple levels of the cortical hierarchy from primary sensory to high-level areas, both of them involve strong top-down effects, and show flexible generalization depending on context.

1.4 A common probabilistic framework for PL and SL

Given the diminishing difference between PL and SL, a parsimonious approach to sensory learning is to define a framework that can seamlessly integrate studies and results in the two domains. A particularly suitable scheme is the probabilistic learning framework that has emerged in the field of machine learning (Ghahramani, 2015), cognitive psychology (Tenenbaum et al., 2011), and neuroscience (Fiser et al., 2010; Knill & Pouget, 2004) over the last two decades. This

framework inherently combines sensory bottom-up and experience-based top-down influences relying on their relative uncertainty to describe information processing in the brain (Fiser et al., 2010; Kersten et al., 2004; Rao et al., 2002). More recent hierarchical extensions of the framework under the name of Hierarchical Bayesian Models (HBM) can potentially capture the full complexity of human learning including high cognitive functions such as abstract concept formation, language acquisition and causal learning (Lake et al., 2015; Tenenbaum et al., 2006).

Our main proposal is that PL and SL should be treated jointly in the framework of HBM, since they are not two separate types of learning, but two extreme testing paradigms of the same complex learning mechanism, in which either more complex structures and context (in case of PL) or the treatment of low level fine sensory features (in case of SL) have been deliberately eliminated (Fig. 1.6A & C). Although there were earlier studies linking the probabilistic framework to either PL (Bejjanki et al., 2011; Michel & Jacobs, 2007) or to SL (Goldwater et al., 2009; Orbán et al., 2008), no studies have explored the benefit of treating PL and SL jointly under the same HBM framework. This is surprising, as the HBM framework inherently fits the overwhelming majority of natural learning situations, where both details of features and the more global structure and context of the sensory information might be relevant for successfully solving the task at hand.

1.4.1 The Bayesian approach

Bayesian approaches usually assume that humans build a mental model of their environment and make (or approximate) optimal (or close to optimal) inferences given their mental model, their current sensory observation, and their uncertainty both in their mental model and in their sensory observation (Griffiths et al., 2010).

A mental model (also referred to as *generative model*) is a representation in the brain that

describes the possible causal processes in the environment that could have generated the sensory input to the senses. Our mental generative model is complex and contains many latent variables that cannot directly be observed (see Fig. 1.5). For example, one can represent the category of a trial in a simple orientation discrimination task (see Fig. 1.1A) with a latent variable (see variable C in Fig. 1.6A). This variable, representing the trial's category, is latent because participants cannot directly observe whether the orientation of a grating stimulus was larger than the reference angle, but instead they have to infer it based on observing the stimulus and the reference angles one by one and compare mentally the two angles to each other.

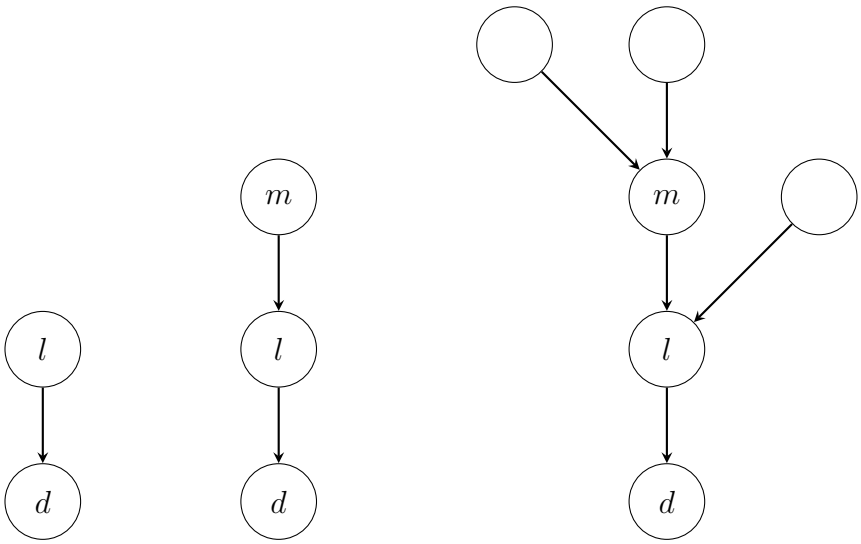


Figure 1.5: Example of abstract generative models with increasing complexity. Left: Simple latent variable model where the unobserved, therefore latent variable, l generates the data, d which is observed. **Middle:** Hierarchical latent variable model where both m and l are unobserved, latent variables forming a hierarchical structure. m generates l which generates the observation/data. **Right:** Illustration how generative models can be extended to capture more complex mental models representing the causal relationships between the variables.

In the Bayesian framework the generative model, derived from the assumed mental model of the observer, is probabilistic, as it represents the uncertainty in some or all of the unobserved, latent variables relevant to the task, including highly abstract and internal-state-dependent contextual latents (Koblinger et al., 2021).

A mental model is also abstract; it does not describe specific sensory events but it represents

possible hidden causal structures that can parsimoniously explain many sensory events jointly, thus allowing for generalization.

Bayesian models provide an optimal framework for inferring latent variables based on the mental generative models given the sensory observations. Assuming a space of possible values that a latent variable can take in a generative model denoted by L, one can update their belief about the value of the latent variable, denoted by $L = l$, after observing data from our senses, denoted by d, using the product and sum rules of probability (see Fig. 1.5 left for the graphical representation of the generative model of this example):

$$\mathcal{P}(l \mid d) = \frac{\mathcal{P}(d \mid l)\mathcal{P}(l)}{\sum_{l' \in L} \mathcal{P}\left(d \mid l'\right) \mathcal{P}\left(l'\right)} \tag{1.1}$$

This update rule is called Bayes' rule and it can be easily extended to multiple latent variables forming a hierarchical structure. Using our previous example in Eq. 1.1, it can also be assumed that the mental model itself is a latent variable and one can infer beliefs in the mental model based on observations. Assume that we have a constrained space of possible mental models denoted by M. One can infer beliefs in a mental model, $M = m$, and its latent variables jointly given our sensory data using the product and sum rules of probability in the same way as in Eq. 1.2 (see Fig. 1.5 middle for the graphical representation of the generative model of this example):

$$\mathcal{P}(m, l \mid d) = \frac{\mathcal{P}(d \mid l)\mathcal{P}(l \mid m)\mathcal{P}(m)}{\sum_{l' \in L} \mathcal{P}\left(d \mid l'\right) \sum_{m' \in M} \mathcal{P}\left(l' \mid m'\right) \mathcal{P}(m')} \tag{1.2}$$

Following the rules of probability in the same way as it was showed in the previous examples in Eqs. 1.1 & 1.2, the Bayesian approach provides an optimal framework to compute the degrees of beliefs in latent variables in any arbitrarily complex model (e.g., see Fig. 1.5 right). Representing the uncertainty in those beliefs associated to the values of the latent vari-

ables in one's mental model allows to make flexible and efficient inferences from scarce data sets (Koblinger et al., 2021; Tenenbaum et al., 2011). However, the data and memory efficient inference is costly in terms of computational complexity. Adding more latent variables will increase the computational complexity of the inference and it becomes intractable for even moderately complex generative models (compare the computational complexity in Eq. 1.1 to Eq. 1.2 where only one more latent variable was added).

Recent research has shown great advances on how Bayesian approaches can be implemented with tractable algorithms. There are two main methods for approximating Bayesian inference with tractable algorithms. First, sampling based methods implement efficient algorithms that can draw samples from probability distributions representing the degrees of beliefs in latent variables in a model (Gilks & Spiegelhalter, 1995). In short, the main advantage of the sampling methods is that it can approximate an integration (or summation), that is too complex to be evaluated, by implementing a summation on only a tractable number of samples drawn from the distribution to be integrated (or summed). It is also possible to draw samples from unnormalized distributions saving the computational expenses of integration (or summation) in the normalization.

Second, the variational inference and expectation propagation methods approximate a true probability distribution with tractable parametric distributions and aim at minimizing the "difference" between the true and the approximate distributions by adjusting the parameters of the approximate distributions (Jordan et al., 1999). More recently, researchers has started to utilize deep neural networks (DNNs) with hidden layers to approximate complex probability distributions (see Ruthotto and Haber, 2021 for more details) and different studies in cognitive science have started to apply these DNNs for capturing complex mental models of humans (Ellis et al., 2018; Nagy et al., 2020). The advances in tractable algorithms approximating Bayesian infer-

ence (Ellis et al., 2018; Lake et al., 2015; Lake et al., 2017) and theories of how the brain can implement such algorithms (Fiser et al., 2010; Orbán et al., 2016; Pouget et al., 2013; Vértes & Sahani, 2018) have made the Bayesian approach an appealing framework for providing a general computational principle of the human brain (see a detailed discussion on the advantages of representing the uncertainty in latent variables Koblinger et al., 2021).

1.4.2 The unifying Hierarchical Bayesian model (HBM)

HBMs, the hierarchical extensions of Bayesian models in perception, can provide a rational, unifying computational framework to explain the flexible generalization effects in PL and SL paradigms. Bayesian inference optimally combines prior knowledge, representing our mental model and beliefs of the task and the stimuli that are independent of the current observation (i.e. top-down influences), with the current sensory observation (i.e. bottom-up influences) while taking into consideration the uncertainty both in the prior knowledge and in the sensory observation. This optimal inference can be applied in hierarchical models (see Eq. 1.2) which can capture the structure of the task and the stimuli, and how it interacts with the sensory representation of the stimulus's feature. In this way, this computational modelling framework can inherently connect learning statistical structures (e.g. temporal or spatial co-occurrences of the stimuli in SL) to learning to detect or discriminate perceptual features (e.g. discriminate orientations from a reference angle in PL).

In order to demonstrate how HBMs can capture PL and SL jointly, let us assume the simplest common generative model across the two types of learning shown in Fig. 1.6C. In the HBM of this example, the observer's perception can be formalized with a probability distribution over

the stimulus (S) given her sensory evidence $\widehat{S}$:

$$\mathcal{P}(S \mid \widehat{S}) = \iint \mathcal{P}(\widehat{S} \mid S, \theta)\ \mathcal{P}(S \mid I)\mathcal{P}(\theta, I)\ \mathrm{d}\theta\, \mathrm{d}I \tag{1.3}$$

where θ, and I denote, the sensory parameters and the structure of the task (c.f. inventory), respectively, and $\mathcal{P}(S \mid \widehat{S})$ captures the observer's belief of the true stimulus given her sensory representation. Since under natural conditions, the observer does not know the structure (I) or the sensory parameters (θ) given the structure, s/he has to learn them jointly:

$$\mathcal{P}(\theta, I \mid \widehat{S}_{1:T}, F_{1:T}) \propto \int \mathcal{P}(\widehat{S}_T \mid S_T, \theta)\ \mathcal{P}(S_T \mid I, F_T)\ \mathrm{d}S_T\ \mathcal{P}(\theta, I \mid \widehat{S}_{1:T-1}, F_{1:T-1}) \tag{1.4}$$

where F denotes the feedback (not shown in the graphical models) and T is the trial number. The three terms on the right side of Eq. 1.4 can be derived from the generative model in Fig. 1.6C and represent the low-level sensory model by $\mathcal{P}(\widehat{S}_T \mid S_T, \theta)$, the high-level representation of the stimulus based on the task structure by $\mathcal{P}(S_T \mid I, F_T)$, and the prior distribution which is the posterior at the previous time step by $\mathcal{P}(\theta, I \mid \widehat{S}_{1:T-1}, F_{1:T-1})$. In this framework, classic PL (Fig. 1.6A) is framed as parameter learning (Michel & Jacobs, 2007), and classic SL (Fig. 1.6B) as structure learning (Orbán et al., 2008). PL without SL emerges when there is no uncertainty in the task structure or the feedback shows the true stimulus, thus the term $\mathcal{P}(S_T \mid I, F_T)$ becomes a Dirac-delta. SL without PL is captured when there is no uncertainty in the sensory process thus the term $\mathcal{P}(\widehat{S}_T \mid S_T, \theta)$ becomes a Dirac-delta. When PL and SL occur jointly, the interaction between the two types of learning can be investigated by using a Joint Statistical Perceptual Learning (SPL) paradigm (Fig. 1.6C) and modelled by Eq. 1.4.

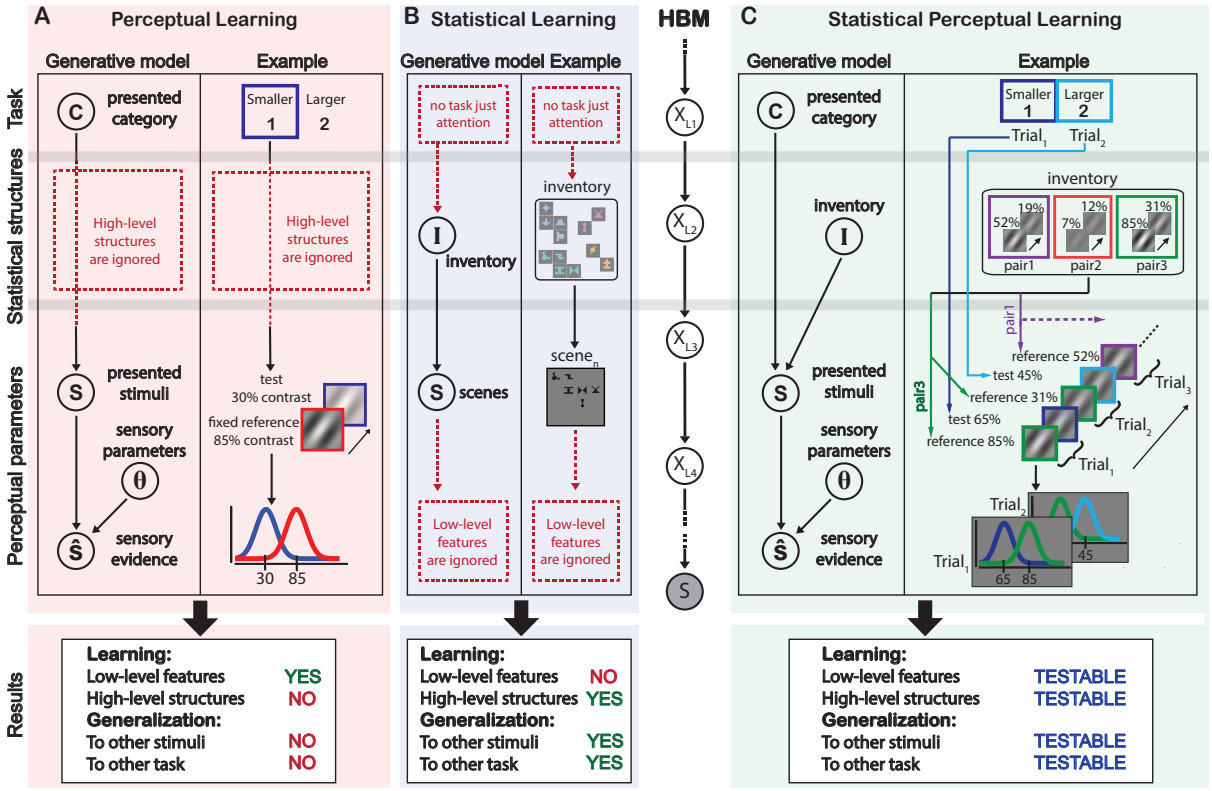


Figure 1.6: Unifying PL and SL in a probabilistic framework. HBM: The scheme of the general Hierarchical Bayesian Model that provides a unified computational framework for classical perceptual (A) and statistical learning paradigms (B), as well as for the combination of the two (C). **A-C:** Probabilistic interpretation of the three paradigms, each with the instantiation of the generative HBM within the given paradigm (left) and one example experiment (right) together with levels not controlled by the paradigm (red dashed rectangles). **Bottom row:** Features of each paradigm and questions that they can address. **A:** PL example of a two alternative forced choice contrast discrimination task. **B:** SL example of visual patterns learning. **C:** Joint Statistical Perceptual Learning (SPL) of contrast discrimination with structured reference stimuli. The reference contrast is not selected randomly but it follows the order defined by sequentially chosen reference contrast-pairs from the inventory. While PL with randomly varying reference contrast levels is excessively hard, I expect that providing a statistical structure to the changes across reference levels (imitating natural conditions) enables and enhances PL. Adapted with permission from J. Fiser and G. Lengyel.

To make this description simple, I intentionally omitted a few important points. First, *structure learning* is often computationally intractable. A common assumption is that the observer only compares a smaller but relevant subset of generative models (structures) or does model averaging (see examples of tractable algorithms for structure learning in Ellis et al., 2018; Kemp and Tenenbaum, 2008; Lake et al., 2015; Orbán et al., 2008). Second, in *structure learning* the

possible parameter values of the structure (θ_I) still need to be marginalized out:

$$\mathcal{P}(I \mid S) \propto \mathcal{P}(S \mid I)\ \mathcal{P}(I) \propto \int \mathcal{P}(S \mid I, \theta_I)\ \mathcal{P}(\theta_I \mid I)\ \mathrm{d}\theta_I\ \mathcal{P}(I) \tag{1.5}$$

Third, Eq. 1.4 assumes a stationary distribution for the sensory parameters (θ), hence the observer estimates a fixed sensory parameter setting. However, it is more realistic to assume that θ changes over time which can be added to the model by assuming a Markov model (e.g. adding transition probability, $\mathcal{P}(\theta_t \mid \theta_{t-1})$ in case of a first order Markov model). Furthermore, in more realistic scenarios the dynamics of the stimuli also needs to be taken into consideration.

Finally, I omitted the response of the observer and how the experimenter can fit her model to the data. In most psychophysical experiments, it can be assumed that the observer makes a response using her posterior distribution over the categories:

$$\mathcal{P}(R \mid \widehat{S}) = f\{\mathcal{P}(C \mid \widehat{S})\} \tag{1.6}$$

where R is the response, C is the category of the stimulus and f denotes a function (e.g. *softmax* function is widely used in decision-making tasks, Neil et al., 2013). Then, the probability of the response given the presented stimulus (S) can be computed by marginalizing out the uncertainty about the unobserved sensory evidence ($\widehat{S}$):

$$\mathcal{P}(R \mid S) = \int \mathcal{P}(R \mid \widehat{S})\, \mathcal{P}(\widehat{S} \mid S)\, \mathrm{d}\widehat{S} \tag{1.7}$$

In this introductory chapter, I provided a high-level description of an HBM that formalizes how PL and SL arise jointly. However, in order to model behaviour in specific PL/SL paradigms, the probability distributions in all of the equations need to be specified assuming

a sensory model representing and updating sensory representations of stimuli' features, and a cognitive model representing and updating the structure in the task and stimuli, and their dynamics. Nevertheless, it is already apparent from the general description above that in the HBM framework, the diverse set of generalization effects in PL and SL is explained by a common computational principle: statistically optimal fusion of prior knowledge with sensory observations. In Chapter 4, I will present a detailed example of how to apply this framework to classical and roving PL paradigms. In short, roving refers to paradigms in which some properties of the PL task are intermixed during training. In the most commonly used roving paradigms participants have to discriminate the stimuli from multiple different references. A number of studies using the rowing paradigm found seemingly contradicting, diverse sets of results that could not be explained jointly by any computational models in PL. Therefore, roving is an ideal paradigm to demonstrate the benefit of treating PL and SL jointly under the same HBM.

1.4.3 Relating HBMs to existing computational models

By explicitly capturing different aspects of the input and the learning task through structured priors, the HBM approach is compatible and includes as a special case the Reweighting models (Dosher & Lu, 2017), two-stage models (Shibata et al., 2014), and the Reverse Hierarchy Theory (Ahissar & Hochstein, 2004) of PL. The main differences between HBMs and the existing computational models in PL can be traced back to the two main characteristics of HBMs; (1) representing the uncertainty in all of the latent variables, and (2) incorporating the top-down influences of higher-level, abstract latent variables by assuming hierarchical structural relationships between the latent variables.

First, Reweighting models (Dosher & Lu, 2017) use feed-forward neural networks which represent the uncertainty only in the latent variable that represents the decision in the PL task.

Therefore, Reweighting models (Dosher & Lu, 2017) can be considered as special cases of an HBM that has uncertainty only in the decision variable, and that does not assume top-down influences.

Second, the two-stage models (Shibata et al., 2014; Watanabe & Sasaki, 2015) do not presume specific latent variables, representations, or learning rules; they only assume that the brain engages in two forms of learning during PL. Those are *feature-based plasticity* during which participants learn to improve the processing and/or the representation of perceptual features, and *task-based* plasticity during which participants learn to improve the processing and/or the representation of the task. Conceptually, this theory reflects a very similar idea to that of HBM; both emphasize the interaction between learning the structure of the task and the perceptual parameters jointly. However, the mechanisms of the two learning types (feature- and task-based plasticity) yet to be explicitly specified for the two-stage model (Watanabe & Sasaki, 2015). Until this happens and it turns out to be significantly different from the formalism of HBM, HBMs can be considered as Bayesian implementations of the two-stage model.

Finally, the Reverse Hierarchy Theory (RHT) (Ahissar & Hochstein, 2004), similarly to the two-stage model, does not deal with specifying representations or learning mechanisms. RHT only presumes that learning in PL follows a gradual top-down direction both in terms of the level of abstraction of the latent variables in a computational model and also in terms of brain areas along the sensory information processing pathways. This means that learning, induced by training on a PL task, will first occur at high-level cognitive areas, then, as the training goes on, learning gradually moves to low-level perceptual areas. HBMs usually assume that the latent variables at all levels in the hierarchy are inferred and updated jointly during learning as the sensory information flows in from the senses. Although this seems to contradict the top-down cascade of learning in RHT at first sight, the joint updating of latent variables at higher- and

lower-levels in particular tasks can easily result in behavior that creates the impression of most learning occurring at higher-levels. Therefore, one can build an HBM that can generate behavior showing a top-down cascade of learning assumed by the RHT.

The HBM approach that I propose here is also compatible with the probabilistic chunk learning models of SL that use the HBM approach and are already known to capture human behavior better than the alternative associative learning and counting models (Austerweil & Griffiths, 2011; Orbán et al., 2008). The HBMs capturing PL and SL jointly is different from the HBMs implemented for SL tasks only (Austerweil & Griffiths, 2011; Orbán et al., 2008) by (1) adding low-level perceptual latent variables, (2) assuming a biologically more realistic model for how the stimuli generate sensory representations, and (3) focusing on how the perceptual parameters are updated during PL task jointly with the parameters capturing the structure of the task and stimuli (see the implemented HBM in Chapter 4). These differences distinguish HBMs from the non-probabilistic chunk learning models of SL as well, (Mareschal & French, 2017; Perruchet, 2019) together with the difference of representing the uncertainty in all of the latent variables in HMBs as opposed to only in the output variable (as in the non-probabilistic chunking models).

To sum up, HBM can accommodate the wide variety of recently established results in the domains of both PL and SL, and facilitates a clearer separation of their causes.

1.4.4 Suggestions for new paradigms investigating the interaction of PL and SL

The integrated viewpoint of HBM also provides a useful guiding principle to identify the kind of experiments that could advance a fuller understanding of the nature of human and animal sensory learning.

The first type of experiments that I propose (Fig. 1.4, Groups A,B) could use multi-element stimuli and semi-relevant cover stories with a PL task to explore how the effect of such sensory and cognitive context could be systematically captured as a consequence of priors acquired earlier by SL. Such an "SPL" experimental setup together with the HBM framework could handle in a coherent manner the phenomena of rowing (Adini et al., 2004; Kuai et al., 2005; Yu et al., 2004), the generalization results of double-training (Wang et al., 2014; Xiao et al., 2008; Xiong et al., 2016), imagination-based learning (Tartaglia et al., 2009c), interaction between orientation detection and categorization (Tan et al., 2019) and perceptual biases due to SL (Luo & Zhao, 2018; Otsuka & Saiki, 2016; Piazza et al., 2018; Yu & Zhao, 2018; Zhao et al., 2013; Zhao et al., 2011; Zhao & Yu, 2016). The SPL paradigm in Fig. 1.6C provides a specific example of how the interaction between SL and PL could be investigated by combining the traditional PL and SL paradigms. SPL can be considered as a rowing paradigm in PL similar to (Kuai et al., 2005). SPL implements a discrimination task with multiple references generated according to a temporal structure. The references form pairs and the elements in the pairs will appear in a particular consecutive order during the task. Previous studies showed that PL with randomly varying reference levels is excessively hard (Adini et al., 2004; Yu et al., 2004), however in SPL the reference is not selected randomly, but it follows the order defined by sequentially chosen reference-pairs from the inventory (Fig. 1.6C), therefore the statistical structure across reference levels (imitating natural conditions) could enable and enhance PL (similar to the results in Kuai et al. (2005)).

The second type of experiments that I propose (Fig. 1.4, Groups C,D) could extend the first one further by using natural scene inputs instead of artificial stimuli and could be applied to explain the high generalization of bird (and other) experts (Devillez et al., 2018; Tanaka et al., 2005), task-structure learning (Kattner et al., 2017) and increased PL performance after video-

game playing (Green et al., 2010a; Green et al., 2010b). Although HBMs in theory can capture the full complexity of human sensory learning in complex, naturalistic, and dynamic environments, tractable algorithms implementing approximate Bayesian inference in such scenarios are still lacking (but see Ellis et al. (2018)). Instead of using entirely natural tasks and stimuli, researchers presently aim at less complex, but still naturalistic and dynamic environments (e.g. (Kwon et al., 2020a)).

1.4.5 Neural implementation for HBMs exploring PL and SL

One of the main obstacles hindering progress in PL and SL research is due to correlating widely different aspects of neural activity with learning (Fig. 1.2, x axis). Although our proposal of introducing HBMs for the computational treatment of learning seems to further complicate this problem, in fact, the probabilistic view offers a unification and clarification on earlier results. As the probabilistic computational framework inherently requires a new type of conversion and approximation from abstract computational descriptions by probability distribution to neural signals (Knill & Pouget, 2004; Pouget et al., 2003), the new representations can provide a principled way to establish a rigorous link between the different types of neural correlates of learning. In particular, sampling-based approximations have been argued to fit well the available neural evidence for perception and learning in the brain (Fiser et al., 2010; Haefner et al., 2016). Sampling-based methods assume that the distribution of latent variables are represented in the dynamics of the neural response directly as samples drawn from the distribution (see Fig. 1.7B). Although there are substantial neural data, analyzing the static (Haefner et al., 2016; Orbán et al., 2016) and dynamic (Echeveste et al., 2020) activity patterns in the primary visual cortex supporting sampling-based models studies involving other sensory domains and brain areas are still lacking.

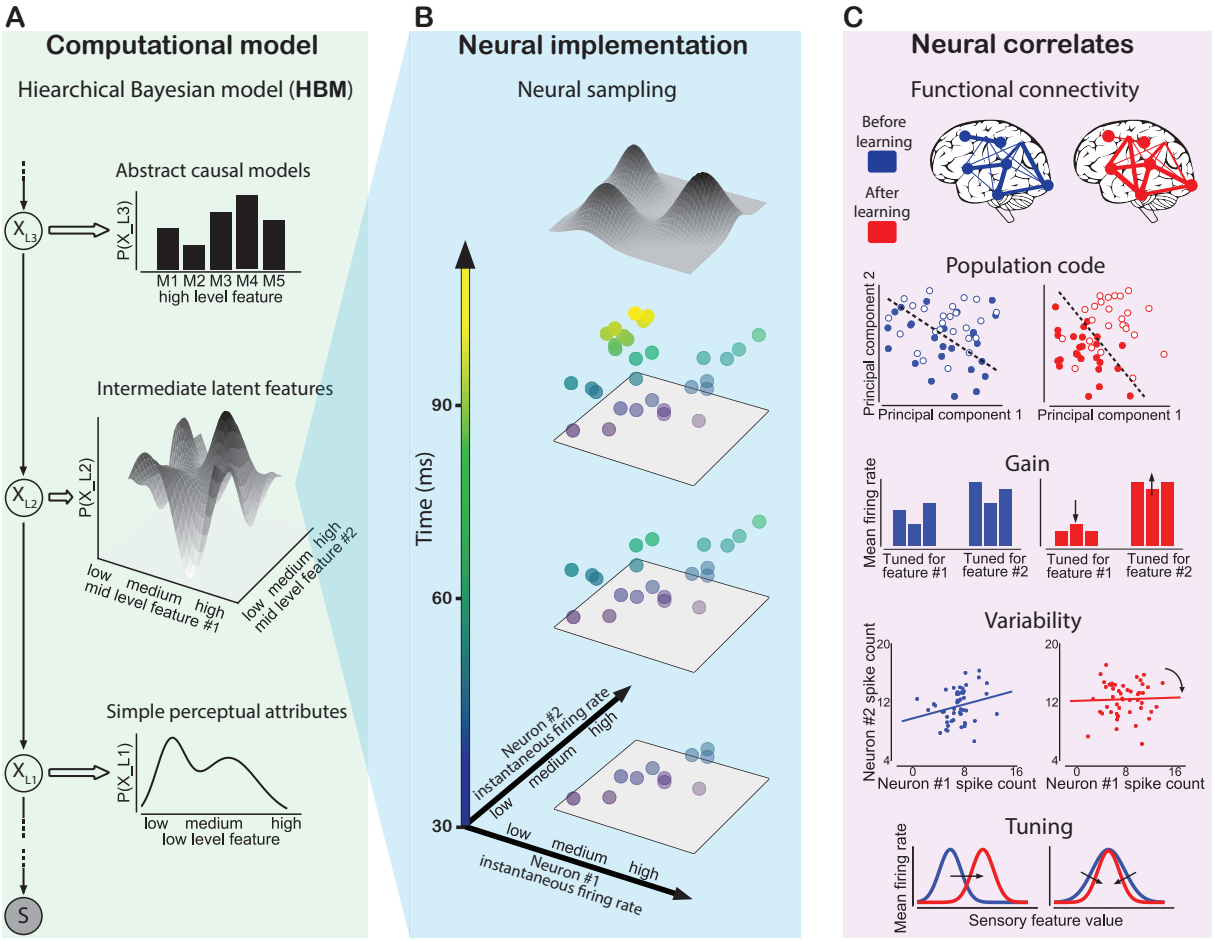


Figure 1.7: Linking the proposed HBM framework for PL and SL to different neural correlates through a probabilistic sampling-based neural implementation. **(A)** In the HBM (left), the stimulus (S) is jointly described by observed and latent features of the environment, which are represented by momentary posterior distributions, $P(X_{Li})$, over possible values of latent variables, X_{Li}, at different levels of abstraction. **(B)** According to the neural sampling hypothesis, covarying neural activities within different cortical areas directly represent the probability distributions over the latent variables of the HBM as samples from that distribution. For each probability distribution (depicted here for latent variables at a middle level of abstraction shown in **(A)**, the individual samples of the joint instantaneous firing rates of neurons at a given time frame (dots) accumulate through time (y axis, also color code of dots), and they jointly approximate the probability distribution of the latent variable (grey 2D distribution on top) with an increasing precision. **(C)** Various previously reported neural correlates of sensory learning that can be potentially derived from the sampling-based probabilistic representation of latent variables. These include shifts and sharpening of tuning curves, decorrelation of neural responses, and changes in gain, population codes (Haefner et al., 2016; Orbán et al., 2016), and, functional connectivity of neural clusters. Adapted with permission from J. Fiser and G. Lengyel.

Various other implementational frameworks can also capture top-down influences of neural signals such as effects of decision making and attention based on recurrency. These include recursive neural network models (Piëch et al., 2013), Predictive Coding (Aitchison & Lengyel,

2017), Probabilistic Population Codes for Bayesian inference making (Pouget et al., 2013) or distributed distributional codes aiming to capture the uncertainty in all latent variables (Vértes & Sahani, 2018). However, sampling-based methods offer a potentially more precise link between computations and various manifestations neural correlations including neural tuning curves, response means and variability, correlations and population sparseness (Orbán et al., 2016) that can likely be recursively extended to higher levels of the hierarchy.

1.5 The goals and the outline of the thesis

The goal of the present thesis is to investigate learning and generalization effects in PL and SL paradigms, and to support the argument that there is a need for a new, unifying framework that can integrate recent results in the two domains of learning and can provide a systematic way to shift PL and SL paradigms closer to each other. In the following paragraphs, I provide the outline of this thesis with brief summaries of the chapters.

This introductory Chapter 1 expands our theoretical paper (Fiser & Lengyel, 2019), in which we proposed for the first time that PL and SL should be treated uniformly and jointly under the HBM framework because this would enable addressing more natural and complex learning problems than before and because, combined with the probabilistic sampling approximation, such a treatment could link more successfully abstract computations of learning with various cortical and subcortical processes. Based on this approach, we proposed a number of new experimental paradigms that can combine the characteristics of current PL and SL paradigms for a more in-depth investigation of human and animal sensory learning and its neural correlates. Hence, this theoretical work sets the stage for the empirical investigations of the following chapters.

Chapter 2 is based on our recent paper (Lengyel & Fiser, 2019) where we investigated

classical PL paradigms looking for general rules predicting learning and generalization in PL. We found that only two general rules were proposed in the PL literature and that in most cases learning and generalization in PL depended both on the structure of the task and the stimuli. The first general rule stated that the amount of learning is proportional to the initial performance (Astle et al., 2013), and the second general rule claimed that the amount of generalization is inversely proportional to the amount of learning (Hussain et al., 2012). In three experiments using contrast and orientation discrimination tasks, we demonstrated that the first rule only reflects the participants' perceptual scaling function and does not show any characteristics of the learning in PL tasks. Regarding the second rule, we found that generalization depended on both the variability and the number of repetitions of the stimuli during training, inducing widely different effects ranging from hindering to enhancing generalization. Thus, there is no evidence for a general rule between the amount of learning and generalization in PL per se as variability depends on the subjective interpretation of the structure of the task and the stimuli which are outside of the domain of classical PL. This further supports the need for a new framework in PL that can account for the diverse results by learning the task structure and the perceptual parameters jointly during PL.

Chapter 3 focuses on two published SL studies (Lengyel et al., 2019) (Lengyel et al., 2021) investigating how learning the statistical structure in the stimuli across scenes in classical SL paradigms influences subsequent perceptual processes in visual search and haptic pulling tasks. Previous studies in SL established that participants build representations that assume the stimuli were caused by abstract latent variables, called "chunks" with a causal structure between the chunks and the stimuli (Mareschal & French, 2017; Orbán et al., 2016; Perruchet, 2019). More recent studies showed that the chunks built during SL influence how we perceive subsequent stimuli (Barakat et al., 2013; Luo & Zhao, 2018; Yu & Zhao, 2018; Zhao & Yu, 2016). Based

on these results, we hypothesized that SL must have an important role in how humans learn to segment their environment into objects. Since objects are the meaningful units/chunks in our sensory environment, they can be considered as the latent variables in the environment generating the sensory input reaching our brain. In two studies, we investigated the "objectness" of the chunks that participants create during classical SL tasks. In the first study, we found that chunks learnt during a classical spatial SL task elicited similar object-based attentional effects as images of true objects with true visual boundaries did (Lengyel et al., 2021). In the second study, we demonstrated a phenomenon of "zero-shot-generalization" namely that the representations of chunks learnt in one modality during an SL task immediately and automatically generalized to another modality without any training in the second modality. Such an instantaneous generalization across modalities indicates that any coherent statistical structure in one domain must immediately be interpreted in context of all other modalities, which is the hallmark of defining representations of abstract objects. These results together suggest that humans built abstract modality independent representations of chunks during SL that serve as perceptual units for interpreting and segmenting subsequent sensory input. These powerful generalization effects point to computations that are in line with the HBM framework for interpreting SL.

In Chapter 4 I investigate the interaction between SL and PL under the same HBM in PL paradigms using roving conditions. Dozens of studies demonstrated that PL is disrupted when observers performed discrimination tasks with multiple references using the same perpetual attribute (Adini et al., 2002; Adini et al., 2004; Amitay et al., 2005; Banai et al., 2010; Cong & Zhang, 2014; Dosher et al., 2020; Kuai et al., 2005; Nahum et al., 2010; Otto et al., 2006; Parkosadze et al., 2008; Tartaglia et al., 2009b; Yu et al., 2004; Zhang et al., 2008). Most explanations suggest that optimizing the performance in discrimination for a reference will interfere with the discrimination performance for another reference, therefore the learning with

multiple references will cause interference effects between the references which will result in no improvement with any of the references (Dosher et al., 2020; Nahum et al., 2010; Tartaglia et al., 2009b; Zhang et al., 2008). However, previous investigations also showed that PL is possible when the reference conditions are in blocked (Adini et al., 2004; Banai et al., 2010; Nahum et al., 2010; Yu et al., 2004; Zhang et al., 2008) or follow a temporal pattern (Cong & Zhang, 2014; Kuai et al., 2005) which suggests that the regularities in the references support PL. This pattern of results can parsimoniously be captured if PL and SL is modelled jointly under the same HBM in which the observer learns the temporal pattern via SL which interacts and supports the PL process. Using simulated data, I show that the interference effect can be reduced between the references and PL emerges if the observer learns the temporal regularity between the reference-conditions. The learnt structure between the reference-conditions helps in disambiguating the references from each other so that the observer can optimize perception separately for the different reference-conditions. Based on these results, I suggest that the new HBM framework can address most of the previously unexplained phenomena in PL using more complex stimuli. Furthermore, since naturalistic scenarios, in the domain of sensory learning, will always involve both SL and PL processes neither of the learning types can be ignored. Thus, the HBM framework, capturing the two forms of learning jointly, provide a parsimonious computation approach for sensory learning.

Chapter 2

General Rules Predicting Performance in PL

2.1 Summary

This chapter investigates the origin of two previously reported general rules of perceptual learning (PL). First, the initial discrimination thresholds and the amount of learning in PL were found to be related through a Weber-like law. Second, it has been claimed that increased training length negatively influences the observer's ability to generalize the obtained knowledge to a new context. To establish the validity of these rules, we conducted a comprehensive investigation using a five-day training protocol during which separate groups of observers performed discrimination around two different reference values of either contrast (73% and 30%, in Study 1, Experiment 1 & 2) or orientation (25° and 0°, in Study 1, Experiment 3). In line with previous research, we found a Weber-like law between initial performance and the amount of learning, regardless of whether the tested attribute was contrast or orientation. However, we also established that this relationship directly reflected observers' perceptual scaling function relating

physical intensities to perceptual magnitudes rather than being a specific add-on, characteristic of learning in a PL paradigm. In addition, we found that with our typical five-day training period, the extent of generalization was proportional to the amount of learning, seemingly contradicting the previously reported diminishing generalization with practice. This result suggests that the negative link between generalization and the length of training found in earlier studies might have been due to overfitting after excessive training and not a necessary feature of learning showing up in all conditions.

These findings support the view that in order to assess the effects of learning and generalization in PL, researchers always have to take into consideration the structure of the task and the regularity in the stimuli as variations in those will fundamentally determine the actual outcome. This is in line with the main proposal of this thesis; PL should be treated in a hierarchical Bayesian framework (HBM, see section 1.4 in the introductory chapter) capturing the structure of the task and the stimuli. The hierarchical Bayesian approach can explain the wide range of generalization results in classical PL paradigms by combining sensory observation (bottom-up influences) with prior beliefs about that task and stimuli (top-down influences) relying on their relative uncertainty. An extended version of the argument and results presented in this chapter was published in Lengyel and Fiser (2019).

2.2 Study 1

2.2.1 Introduction

In the last decades, numerous factors were identified that influence observers' ability to improve their performance in low-level perceptual tasks after extensive practice, a process termed perceptual learning (PL). Among these factors are feedback (Aberg & Herzog, 2012; Herzog &

Manfred, 1997; Petrov et al., 2006; Seitz & Watanabe, 2003), experimental design (Adini et al., 2004; Kuai et al., 2005; Yu et al., 2004), the nature of the contextual elements around the target (Adini et al., 2002; Manassi et al., 2012), or more broadly, the structure and the variability of the stimuli and the task Cohen et al., 2013; Hussain et al., 2012; Kuai et al., 2005. More recently, the generalization of learning in perceptual tasks also came under investigation, and once again, researchers identified a great number of factors that determine the extent of generalization. Among others, task difficulty (Ahissar & Hochstein, 1997), precision (Jeter et al., 2009), stimulus variability (Hussain et al., 2012), training length (Ahissar & Hochstein, 1997; Jeter et al., 2010), additional tasks and stimuli (Hung & Seitz, 2014; Wang et al., 2014; Xiao et al., 2008; Zhang et al., 2010), and statistical structure of the task and stimuli (Cohen et al., 2013) have an effect on the level of generalization. Although these studies broadened our understanding of the underlying processes of perceptual learning only few of them can provide support for general rules that could predict perceptual learning performance under different conditions (Ahissar & Hochstein, 1997; Astle et al., 2013; Hussain et al., 2012; Jeter et al., 2010). The present study focuses on two previously investigated more universal rules that were suggested to predict performance in perceptual learning paradigms in general: the link between initial performance and the magnitude of perceptual learning (Astle et al., 2013), and the connection between the amount of learning and the extent of generalization (Hussain et al., 2012; Jeter et al., 2010).

2.2.1.1 The relationship between initial performance & learning

Several studies reported that the amount of learning in perceptual tasks (as defined by the improvement in performance from the first day to the last one) can be predicted from the initial performance (Aberg & Herzog, 2009; Astle et al., 2013; Fahle, 1997; Fahle & Henke-Fahle,

1996; Polat et al., 2012; Yehezkel et al., 2016). The earlier examples of these studies used one-interval 2-AFC hyperacuity tasks (Vernier, curvature, and orientation discrimination tasks). These studies found that the better observers' initial performance was the smaller they improved on the task (Fahle, 1997; Fahle & Henke-Fahle, 1996). However, a more recent study by Astle and colleagues (2013) investigated this relationship in more depth and argued for a specific, Weber-like relationship (Fechner, 1999; Weber, 1834) between the initial performance and the magnitude of learning which could reflect a general rule predicting learning in PL paradigms.

In their study, monocular Vernier acuity was measured at various eccentricities in a one-interval 2-AFC task after observers were trained at both 5° and 15° off the central fixation. The authors found that the initial discrimination thresholds, on average, were higher at 15° eccentricity than at 5°. Critically, the amount of improvement on the Vernier acuity task (measured as a difference of the first and the final day's Vernier discrimination threshold in arcsec) was proportional to the initial discrimination thresholds (Astle et al., 2013). In addition, when they equated the observers' initial thresholds at the various eccentricities in the acuity task by spatially scaling the Vernier lines or by visual crowding, the magnitude of learning became equal at the different eccentricities. Thus, regardless of what constraint limited the initial discrimination thresholds prior to training (retinal location, stimulus size, or crowding), the amount of absolute learning seemed to be proportional to the initial threshold level. To further specify this claim, Astle et al. (2013) expressed the relative learning as the observers' ratio of the first and the final day's thresholds (measured in Vernier discrimination threshold in arcsec divided by the line length also in arcsec) and showed that this relative learning did not correlate with the initial thresholds, but it was constant across different levels of initial thresholds. Because this pattern is captured by Weber's law, Astle and colleagues posited that "...perceptual learning also obeys a similar Weber-like law..." and that "...the finding that improvements in normal subjects are

tied to their initial threshold in a lawful way, analogous to Weber's law, suggests that the same factors that impose limits on a visual threshold also constrain the amount an organism can learn on a visual task...." (Astle et al., 2013, pp. 4 and 7).

Astle and colleagues' results (a Weber-like law for absolute learning leading to no correlation in terms of relative learning) are in contrast with those of earlier studies that used the same measure of relative learning, but did report a positive correlation between the relative learning and initial performance in Vernier (Fahle & Henke-Fahle, 1996) and bisection acuity tasks (Aberg & Herzog, 2009). The positive correlations found in those studies means that the amount of absolute learning measured in those experiments was a progressively increasing fraction of the initial discrimination thresholds, implying a power-like law (Stevens, 1957) rather than a Weber-like law.

The discrepancy between the results of the above studies can be tracked back to the issue of whether the relationship between learning and initial threshold is influenced by something else beyond the observers' perceptual scaling function. In psychophysics, the observer's perceptual scaling function represents how physical stimulus intensities are related to perceived magnitudes. Assuming that the discrimination threshold is limited by constant and independent Gaussian noise in accordance with signal-detection theory in its most basic form (Green & Swets, 1966), the perceptual scaling function can be estimated by measuring the observer's discrimination thresholds at different physical stimulus intensities. As the discrimination thresholds represent the lowest increment in the stimulus intensity that the observer can still perceive (at a certain performance level), the scaling function approximates how the observer maps the physical stimulus onto her internal perceptual space (see Figure 2.1, top). In this chapter, we argue that the proportional Weber-like relationship between initial discrimin-

ation thresholds and the amount of learning (Figure 2.1, bottom) emerges in perceptual learning tasks when (a) observers improve by the same amount at different region in their internal perceptual intensity space, and (b) the perceptual scaling function between the perceptual and physical spaces does not change during learning. In this case, the amount of learning will depend only on the same perceptual scaling function of the observer that also determines her initial discrimination threshold prior to learning. Consequently, there will be a proportional relationship between initial threshold and the amount of learning. In contrast, power-like law (or any not proportional functional relationship) between initial discrimination thresholds and learning would emerge only when, in addition to the perceptual scaling function, either a change in the perceptual scaling due to learning or some other additional learning-specific factors affect perceptual learning.

Figure 2.1 shows two simple examples demonstrating the argument above with the two typical perceptual scaling functions found in human perception. In the top plots on the left, the hypothetical observer has a Weber-like perceptual scaling function that transforms physical intensity to perceived magnitude. In the plots on the right, the hypothetical observer has a power-like perceptual scaling function. Initial discrimination thresholds (i.e., the initial just noticeable differences, JNDs) are the smallest step sizes on the stimulus intensity space (x axis) that have a corresponding one unit change on the observers' perceptual space (y axis). In Figure 2.1 we show the initial thresholds for 30 and 59 base-intensities denoted by $\Delta S_{30}(\text{pre})$ and $\Delta S_{59}(\text{pre})$ respectively. The amount of perceptual learning for the two base-intensities is the difference between discrimination thresholds before and after the practice sessions:

$$PL_{30} = \Delta S_{30}(\text{pre}) - \Delta S_{30}(\text{post}) \text{ and } PL_{59} = \Delta S_{59}(\text{pre}) - \Delta S_{59}(\text{post})$$

Using both the Weber-like and the power-like perceptual scaling functions, the same amounts

of improvement in the perceptual intensity space at different base-intensities (colored ranges on y axis) will lead to different amounts of improvement in the stimulus intensity space (colored ranges on x axis) depending only on the shape of the perceptual scaling function linking physical and perceptual intensities. Therefore, both the initial thresholds (pre) and the amounts of learning (pre - post) follow the same function, the observers' perceptual scaling function. This condition will automatically lead to changes in the amount of learning that is proportional to the initial thresholds:

$$\frac{\Delta S_{30}(\text{pre})}{\Delta S_{59}(\text{pre})} = \frac{\Delta S_{30}(\text{pre}) - \Delta S_{30}(\text{post})}{\Delta S_{59}(\text{pre}) - \Delta S_{59}(\text{post})}$$

This theory is only true if (a) the shape of the perceptual scaling function does not change during learning, and (b) the same amounts of improvement occur on the internal perceptual space at the different base-intensities (such as the red and green shaded areas on the y axis). These assumptions are most probably met in classical PL paradigms for the following two reasons. First, the appearance of the investigated perceptual features in the artificial stimuli are not very different from the appearance of those features in natural scenes, thus participants' perceptual model of representing those low-level features (i.e. their perceptual scaling function) should remain stable during PL. Second, in most PL tasks participants receive the same number of training trials at the different base-intensities, therefore the same amount of learning is expected at those base-intensities. However, this linear proportionality vanishes if either the observers' functional mapping from physical to perceptual intensities is modulated during learning or the amounts of improvement on the internal perceptual space are different at different base intensities.

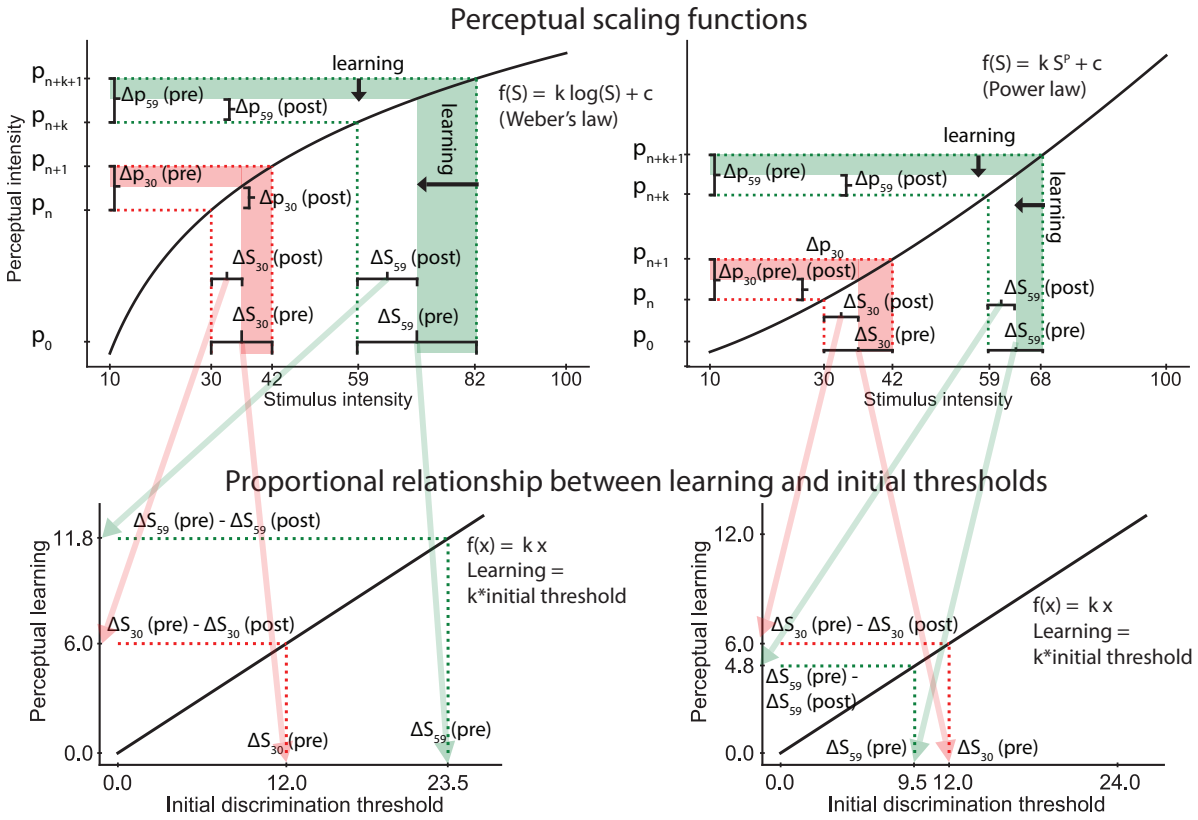


Figure 2.1: The relationship between initial discrimination thresholds and the amount of learning, and how this relationship is related to observers' perceptual scaling function linking physical and perceived intensities. Top: Two perceptual scaling functions: the Weber's law (Left) and the Power law (Right). Physical intensities on the x axis show a hypothetical scale of an attribute from 10 to 100, while the perceptual intensities on the y axis scale from the absolute threshold ($P0$). The scale on the y axis depends on the function, $F(\mathrm{S})$ that maps the physical magnitudes onto the perceptual intensities. Two initial discrimination threshold levels at two base-intensities are shown, at 30, $\Delta \mathrm{S}_{30}(\mathrm{pre})$, large black brackets between the red dotted lines, and at 59, $\Delta \mathrm{S}_{59}(\mathrm{pre})$, large black brackets between the green dotted lines. These initial discrimination thresholds, $\Delta \mathrm{p}_{30}(\mathrm{pre})$ and $\Delta \mathrm{p}_{59}(\mathrm{pre})$ are the smallest perceivable changes at the measured base-intensities. If (1) the same amounts of learning measured on the perceptual sensitivity intensity space at the two base-intensities and (2) the equal amounts of improvement on the perceptual scale will be transformed back with the inverse of the same perceptual scaling function into changes in the stimulus intensity (colored changes on the x axis) then, the amounts of learning at different base-intensities (e.g., $\Delta \mathrm{S}_{30}(\mathrm{pre}) - \Delta \mathrm{S}_{30}(\mathrm{post})$) will follow the same perceptual scaling function that determined the initial discrimination thresholds (e.g., $\Delta \mathrm{S}_{30}(\mathrm{pre})$) prior to learning. Consequently, proportional relationship between the initial discrimination thresholds and the amount of learning at the two base intensities emerges: $\Delta \mathrm{S}_{30}(\mathrm{pre}) \,/\, \Delta \mathrm{S}_{59}(\mathrm{pre}) = [\Delta \mathrm{S}_{30}(\mathrm{pre}) - \Delta \mathrm{S}_{30}(\mathrm{post})]/[\Delta \mathrm{S}_{59}(\mathrm{pre}) - \Delta \mathrm{S}_{59}(\mathrm{post})]$ **Bottom**: The proportional relationship between initial thresholds (IT) and perceptual learning (PL). Initial discrimination thresholds, $\Delta \mathrm{S}(\mathrm{pre})$ are shown on the x axis, while the amount of learning, $\Delta \mathrm{S}(\mathrm{pre}) - \Delta \mathrm{S}(\mathrm{post})$ on the y axis. The dotted red and green lines represent the corresponding initial discrimination threshold levels and the amount of learning at 30 and 59 stimulus intensities derived from the top panels. The green and red arrows show the relationship between the top and the bottom figures for the two stimulus intensities. Regardless of the exact perceptual scaling function the relationship between learning and initial thresholds remains proportional: $PL = k\ IT$, with k as a scaling constant. Adapted with permission from J. Fiser and G. Lengyel.

Using the above observations, the difference between the findings of Astle et al. (2013) and Aberg and Herzog (2009) and Fahle and Henke-Fahle (1996) can be captured as follows. Astle et al.'s result can be explained by assuming that (a) observers improve by the same amount at different base-intensities in their internal perceptual space, and (b) their perceptual scaling function does not change during perceptual learning. In this case, the amount of learning depends only on the observers' perceptual scaling function, without assuming any learning-specific extra Weber-like process they posit in their paper. In contrast, assuming a change in the scaling function during learning and/or different amounts of learning at different base-intensities in the internal perceptual space would distort the Weber-like proportionality between initial threshold and learning, confirming Aberg and Herzog's and Fahle and Henke-Fahle's results. In this case, the amount of learning cannot be predicted from the observers' scaling function suggesting that other learning-specific processes are involved during perceptual learning. The first goal of the present study was to investigate which of these two scenarios hold in general during perceptual learning.

2.2.1.2 The relationship between learning & generalization

Traditionally, the specificity of the acquired ability has been a defining hallmark of perceptual learning (Crist et al., 1997; Fahle, 1997; Karni & Sagi, 1991; Schoups et al., 1995). According to this view, whatever improved ability observers develop after extensive training within the context of low-level visual discrimination tasks, this new skill remains available only within the close context of the original setup including the stimulus identity and the location of training in the retinal space. However, recent studies finding substantive transfer of learning under various conditions strongly challenge this notion (Ahissar & Hochstein, 2004; Wang et al., 2014; Zhang et al., 2010).

Investigating the relationship between the amount of learning and generalization involves an inherent ambiguity at the conceptual level. Intuitively, generalization and learning should go hand in hand: more learning means more knowledge about the state of the world and hence, more potential for using the newly learned competence in different contexts. However, it is well-known in the field of machine learning that too much repetitive learning can result in a representation (an internal model) that is overly specific to the trained features and the circumstances of the training, a phenomenon called overfitting (Hastie et al., 2013; Murphy, 2012). In perceptual learning, learning can be defined as the improvement in task performance in a context-specific manner (in the trained condition), while generalization is the improvement in task performance in a context-free manner (in an untrained condition). Overfitting is related to the difference between the two. Excessive training in perceptual learning could cause overfitting, which could lead to a little extra learning, but it also substantially decreases generalization. Indeed, several behavioral studies in the domain of perceptual learning confirmed this conjecture (Hussain et al., 2012; Jeter et al., 2010).

Thus, the relationship between perceptual learning and generalization can depend intricately on two separate components: while the specific features of the learning process, such as the selected task or the training stimuli, could lead to more specific or more generalizable knowledge, overtraining itself can shift performance from flexible to specific. Since overfitting is a general rule in computational learning theories, we were interested in exploring the first component, whether more perceptual learning produces more generalizable knowledge before the effect of overfitting emerges. If more training in various perceptual tasks leads to more learning due to an improved internal model incorporating the actual experience in the observer's world model, proportionally more generalization is predicted before overfitting occurs. However, if more training results in more learning due to focusing only on specific features of the task/stimuli

without viable integration of this knowledge to other aspects of the observer's internal model, learning is expected to be proportionally more specific to the features of the training examples even before overtraining happens. Previous studies modulated the extent of training (Hussain et al., 2012; Jeter et al., 2010) which although influenced the amount of learning, also increased the amount of training data from the same kind rather than providing more new information with the training data which increases the chance for overfitting (Hastie et al., 2013; Murphy, 2012). We used a 5-day long fixed length training protocol to control for the effect of overfitting and measured the individual differences in the amount of learning and generalization in two widely used perceptual learning paradigms (contrast and orientation discrimination tasks). This setup allowed us to pursue the second goal of the present study, to determine whether the extent of generalization is proportional to the amount of learning.

2.2.1.3 Overview of the study

In three experiments, we measured contrast and orientation discrimination thresholds and the amount of learning at two different stimulus intensities (at 73% and 30% contrast, and at 25° and 0° orientation in separate temporal 2-AFC discrimination tasks) and found that the amount of perceptual learning was proportional to the initial performance. Furthermore, we showed that this specific relationship between initial performance and learning mainly reflected the observers' internal perceptual scaling function from physical to perceptual intensities. Our results also revealed a positive link between the amount of learning and generalization: more learning led to proportionally more generalization. We interpreted the relationship between these results and earlier reports in the light of differences in methodological and conceptual characteristics of perceptual learning paradigms.

2.2.2 Methods

Participants

One hundred and twenty naive observers gave informed consent prior to participation in the experiment. Nineteen observers took part in Exp. 1, the within-subject contrast discrimination experiment. In Exp. 2, 25 observers completed the 30% reference condition while another 24 observers the 73% reference contrast condition. In the orientation discrimination experiments (Exp. 3), 15 and 15 observers participated in the 0-degree and 25-degree reference conditions, and another 11 and 11 observers completed the 15-degree and the 45-degree reference orientation conditions, respectively. None of the observers had any previous experience with a psychophysical experiment. All participants had normal or corrected-to-normal vision. The experimental protocols were approved by the Ethics Committee for Hungarian Psychological Research.

Apparatus & stimuli

We used Matlab Psychtoolbox 3 (Brainard, 1997; Pelli, 1997) to generate the stimuli on a 21-in Samsung Syncmaster 1100 DF color monitor (1024 x 768, 85 Hz frame rate, 0.2 mm pixel pitch). The mean luminance was 60 cd/m2. The monitor was calibrated, and the luminance was linearized by X-Rite i1Profiler device and software. The participants viewed the stimuli binocularly at the fovea in a dimly lit room. In both paradigms, the stimuli were Gabor patches defined by Gaussian enveloped sinusoidal gratings with (spatial frequency of 6 cycles/degree (SD: 0.17°), contrast 0.47 in the orientation discrimination task, orientation 36° in the contrast discrimination tasks, and phase randomized for every stimulus presentation in the orientation discrimination task). The Gabor patches were presented on a background at mean luminance. The stimuli were viewed from 2 meters through a circular aperture (diameter 17°) of a black

piece of cardboard that covered the entire monitor screen. The whole cardboard and the viewing area in front of the observer was further covered by a black curtain with a circular aperture (diameter 17°). This setup was used to prevent observers from using the edges of the display in the orientation discrimination task.

Procedure

Investigating initial thresholds & the amount of learning: We conducted perceptual learning experiments using two attributes, contrast and orientation, and we measured discrimination thresholds from a reference value. To test whether perceptual learning was proportional to the initial performance due to the internal perceptual scaling of the participants, we used two experimental conditions in all experiments. In the two conditions, observers were trained with two different stimulus intensities that were known to elicit different initial discrimination threshold levels according to previous studies measuring the perceptual scaling functions of the observers between physical and perceptual magnitudes. In Experiments 1 & 2 the two conditions were distinguished by the reference contrast (30% vs. 73%) at which the observers were trained. Based on previous studies (Burton, 1981; Legge, 1981), we expected significantly higher initial discrimination thresholds at 73% contrast. In Experiment 3, observers were trained at reference orientation of 0° vs. 25°. Again, since earlier studies reported the lowest discrimination threshold at the cardinal orientations (Mach, 1861; Mansfield, 1974; Mikellidou et al., 2015; Orban et al., 1984; Regan & Price, 1986), we expected higher initial discrimination thresholds at 25°. Once the initial discrimination thresholds from the reference values were measured, we assessed the amount of perceptual improvement in each of the conditions and checked whether they showed proportionally more learning in the conditions with higher vs. lower initial discrimination threshold levels.

There is a trade-off in benefits when using within- vs. between-subject designs in perceptual learning tasks. On the one hand, a related-samples statistical analysis in a within-subject design is more sensitive, and therefore, it can reveal a relationship between initial performance and learning even if the individual differences in perceptual performances are large. On the other hand, a within-subject design is potentially prone to uncontrolled generalization between the conditions, which can bias the comparison between the low and high initial performance conditions. To control for this problem, participants in Exp. 1 trained with both reference contrast conditions, 30% and 73% (within-subject design, Fig. 2.2B) similarly to Astle et al.'s (2013) study, in which participants trained at both 5° and 15° eccentricities. Meanwhile, in Exps. 2 and 3, two separate groups of observers were trained with either 30% or 73% reference contrasts in the contrast discrimination task, and with either 0° or 25° in the orientation discrimination tasks (between-subject designs, Fig. 2.2B).

Investigating generalization & the amount of learning: Generalization was quantified by measuring discrimination thresholds at an untrained reference contrast or orientation after finishing the training sessions. In Exp. 1 after training with reference contrast at 30% and 73%, generalization was assessed by measuring discrimination threshold at the untrained 47% contrast. In Exp. 2 the group that practiced with reference contrast at 30%, generalization was tested at both 47% and 73% contrast levels, and the group that trained with reference contrast 73% the transfer of learning was tested at contrast of 30% and 47%. In the orientation discrimination task, generalization in the group that trained at reference orientation 0° was measured at 25°, and for the group that practiced with reference orientation 25° it was assessed at 0°. We tested whether more learning caused proportionally more generalization by assessing the within-condition correlation between individual differences in learning and in generalization. Due to very small inter-subject variability in perceptual performances at 0° reference orienta-

tion, two additional groups of participants completed the very same experiment, but one group trained with reference orientation 15° and generalization was measured at 45°, and the other group trained with reference orientation 45° and generalization was assessed at 15°. In these groups we had sufficiently large inter-subject variability to test our question about generalization.

General procedure: Contrast and orientation discrimination thresholds were measured with a temporal two-alternative forced choice (2-AFC), 1-up-3-down staircase procedure. In each trial, a fixation point was first flashed for 200 ms and disappeared 200 ms before the onset of the first stimulus interval. Next, the reference (contrast or orientation) and test patch were presented after each other for 91 ms each in a random order. The reference and the test patch were separated by a 600 ms interstimulus interval (Fig. 2.2A). In all experiments, observers trained for 5 consecutive days completing one session per day (Fig. 2.2B). In each trial, the observer had to judge whether the stimulus has a more clockwise orientation (in the orientation discrimination task) or a higher contrast (in the contrast discrimination task) in the first or the second stimulus interval. Observers responded by pressing "1" or "2" keys on the keyboard. In all tasks, there was an auditory feedback marking incorrect responses.

The staircase during the experiments followed the 3-down-1-up rule with a step size of 0.05 log units, which converged to 79.4 % correct responses (Levitt, 1971). The initial difference values between the reference and the test for the very first staircase were $\Delta 8\%$, $\Delta 12\%$ for reference contrasts 30% and 73%, $\Delta 8^\circ$ and $\Delta 14^\circ$ for reference orientation 0° and 25°, and $\Delta 8^\circ$ and $\Delta 18^\circ$ for reference orientation 15° and 45°, respectively. The initial differences were determined based on the mean initial discrimination thresholds of the observers in our pilot perceptual learning experiments using the same procedure to approximate contrast and orientation discrimination thresholds. After completing the first staircase, the initial values for the following

staircases were adjusted separately for each observer by taking the observer's average performance in the previous staircase in the same condition and multiplying it by two. Each staircase contained four practice and six experimental reversals. The observer's threshold was defined as the geometric mean of the experimental reversals. Observers completed 5-5 staircase blocks in each reference value condition in the pre-and posttest sessions and 10 staircase blocks with the practiced reference value during each training session. Previous results using simulations suggested that the adaptive method described above should reveal observers' thresholds at 79.4% performance level quite accurately (García-Pérez, 1998). However, in those simulations attentional lapse rates were assumed to be zero and estimating discrimination thresholds based on the stimulus strengths at reversal points could be confounded by attentional lapses (Solomon & Tyler, 2017). Although theoretical work and simulations showed that the 3-down-1-up staircase is robust to the initial attentional lapses (Karmali et al., 2016), lapses are not necessarily limited to the initial trials in novice observers. In order to confirm that the measured decrease in thresholds after practice using the 3-down-1-up staircase method was not just due to the decrease in attentional lapses of our participants, we estimated the lapse rates and the thresholds for each observer by fitting psychometric curves to their performance at pre- and posttests (see the detailed methods in the Appendix A). We found that the thresholds decreased significantly after the training in all experimental conditions (see Fig. A.1) even when we controlled for the decrease in lapse rates (see Fig. A.2). Furthermore, the decrease in thresholds due to learning estimated by the best-fitting threshold parameter of the participants' psychometric curves positively correlated with the estimated decrease in thresholds using the adaptive staircase method (see Figs. A.1 & A.2). This suggests that perceptual learning measured by the thresholds at pre- and posttests using the staircase method reveals perceptual and not just attentional improvement.

Analysis

Exclusion criteria: We excluded outlier participants from the analysis if their performance (in initial thresholds or learning) was more than 2 standard deviations away from the group average. Using this criterion, we excluded two subjects from Exp. 1 because one of them had large negative learning in the reference contrast 30%, and the other one had large negative learning in the reference contrast 70% conditions. We excluded one-one subjects from each of the conditions in Exp. 2 for the same reason: both participants showed large negative learning. There were no outliers in the orientation discrimination task, thus we did not exclude anyone from the analysis in Exp. 3.

Assessing learning: To measure the amount of perceptual learning we used three types of learning scores.

(1) Absolute learning computed as:

$$PL^{abs} = PRE - POST\ (thresholds)$$

(2) Relative learning computed as:

$$PL^{rel} = \frac{PRE}{POST}\ (thresholds)$$

(3) Predicted learning computed as (see Results for more details):

$$PL^{pred} = \frac{PRE_{@LowStimIntensity}}{PRE_{@HighStimIntensity}}\ PL^{abs}$$

Assessing generalization: The amount of generalization was assessed with two metrics.

(1) Absolute generalization:

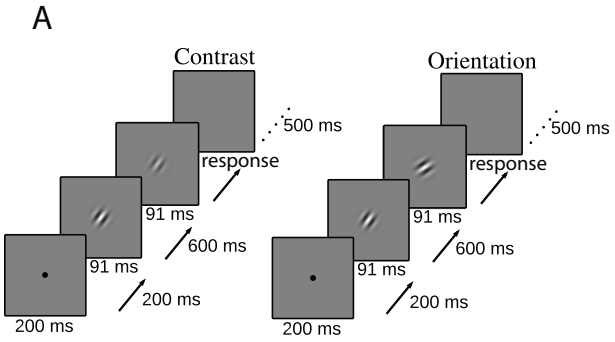
$$PRE - POST \text{ (thresholds at the untrained reference values)}$$

(2) Relative generalization:

$$\frac{Generalization}{Learning} = \frac{PRE - POST \text{ (thresholds at the untrained reference values)}}{PRE - POST \text{ (thresholds at the trained reference values)}}$$

Comparing group means: In our analysis, we needed to evaluate the probability of no difference between two groups' scores, and the probability of certain scores not being different from zero. However, frequentist hypothesis testing cannot confirm the null hypothesis due to its design (Morey & Rouder, 2011; Streiner, 2003).Therefore, we ran independent or paired samples t-tests and also non-overlapping hypotheses (NOH) Bayes factor (BF) analysis for independent or related samples (Morey & Rouder, 2011) to compare the different conditions in the experiments. We computed the Non-overlapping hypotheses (NOH) Bayes factors (BF) (Morey & Rouder, 2011; Rouder et al., 2009) to obtain the level of confidence in concluding no difference between certain learning scores (see Results for the specific comparisons). The NOH BF represents the probability of "there is no or negligible difference between the conditions" divided by the probability of "there is difference between the conditions". Therefore, BFs larger than one indicate how many times more probable the "no or negligible difference" than the "existence of a difference" between the conditions is. In the NOH BF analysis, the null hypothesis states that the effect size is within the range of -0.2 and 0.2, whereas the alternative hypothesis is that the effect size is outside that range. The range of the null hypothesis was chosen following the guidelines of Cohen (2013) and Morey and Rouder (2011) that below 0.2 the effect is negligible. We used a scaling factor equal to one in the scaled Cauchy prior.

Analyzing the variability within conditions: Inter-subject variability was analyzed with Pearson and partial correlation. We applied partial correlation between the amount of learning and the extent of generalization while controlling for the initial threshold levels. The partial correlation coefficient reveals the correlation between the residuals of the linear regressions predicting separately generalization and learning from initial thresholds. If the deviations (residuals) from the predicted generalization and from the predicted learning (using the initial discrimination thresholds as predictor in both cases) correlate, it also indicates a relationship between generalization and learning alone without the influence of the initial thresholds. The partial correlation coefficient between X (independent variable) and Y (dependent variable) while controlling for Z (dependent variable) and the standardized regression coefficients of X in a multiple linear regression predicting Y with both X and Z as predictors gives the same amount of information and p values. Therefore, computing partial correlation between learning and generalization while controlling for initial threshold levels is equivalent to using multiple linear regression to predict the extent of generalization using the initial threshold levels and the learning scores as independent variables.



B

	Day1 - Pretest	Day2-4 - Training	Day5 - Posttest
Exp.1 Contrast	Assesing thresholds at reference 30 AND 73	Training with reference 30 AND 73	Assesing thresholds at reference 30 AND 73
Exp.2 Contrast	Assesing thresholds at reference 30 AND 73	Training with reference 30 OR 73	Assesing thresholds at reference 30 AND 73
Exp.3 Orientation	Assesing thresholds at reference 0° AND 25°	Training with reference 0° OR 25°	Assesing thresholds at reference 0° AND 25°

Figure 2.2: **A**. Contrast and orientation discrimination tasks. **B**. Training protocol. Adapted with permission from J. Fiser and G. Lengyel.

2.2.3 Results & Discussion

2.2.3.1 The ratio of initial performance & learning

We confirmed that the chosen reference values, indeed, led to groups with higher initial discrimination thresholds at high reference values (73% in the contrast and 25° in the orientation discrimination tasks) than at low reference values (30% contrast and 0° orientation). Specifically, we found significant differences between initial discrimination threshold levels in all experiments: in Exp. 1 (t_{16}=7.847, P<0.001, d=1.889), in Exp. 2 (t_{45}=5.852, P<0.001, d=1.664) and in Exp. 3 (t_{28}=6.718, P<0.001, d=2.539) (Fig. 2.3, subpanels A in all panels). This finding means that observers had larger discrimination thresholds around 73% contrast than around 30%, which is in line with previous findings showing a near logarithmic perceptual scaling function from physical to perceived contrast intensity (Burton, 1981; Legge, 1981). In case of the orientation discrimination task, we also found the expected advantage in the discrimination sensitivity at the cardinal orientation (Mansfield, 1974; Mikellidou et al., 2015; Regan & Price, 1986), that is a larger discrimination threshold around 25° than around 0°.

There was significant perceptual learning in all conditions (ps<0.005), although not every observer improved after the training (Fig. 2.3, B and C subpanels in all panels). Perceptual learning was stronger in conditions with higher initial threshold levels (Exp. 1: t_{16}=2.567, P=0.021, d=0.693; Exp. 2: t_{45}=2.126, P=0.039, d=0.631; Exp. 3: t_{28}=4.498, P<0.001, d=1.700, Fig. 2.3, B subpanels in all panels). The ratio of the initial threshold levels and the ratio of the amount of learning in the two conditions were almost the same in all experiments.

Exp. 1: $\frac{IT_{Con30}}{IT_{Con73}} = 0.56 \approx \frac{PL_{Con30}}{PL_{Con73}} = 0.51$

Exp. 2: $\frac{IT_{Con30}}{IT_{Con73}} = 0.53 \approx \frac{PL_{Con30}}{PL_{Con73}} = 0.49$

Exp. 3: $\frac{IT_{Ori0}}{IT_{Ori25}} = 0.41 \approx \frac{PL_{Ori0}}{PL_{Ori25}} = 0.35$

where IT and PL represent initial thresholds and perceptual learning, respectively. While

these results suggest that the amount of learning is roughly proportional to the initial threshold levels, in the next section we perform a statistical test of the exact proportional relationship and show that it reflects the observers' perceptual scaling function which links physical intensity to perceptual magnitude.

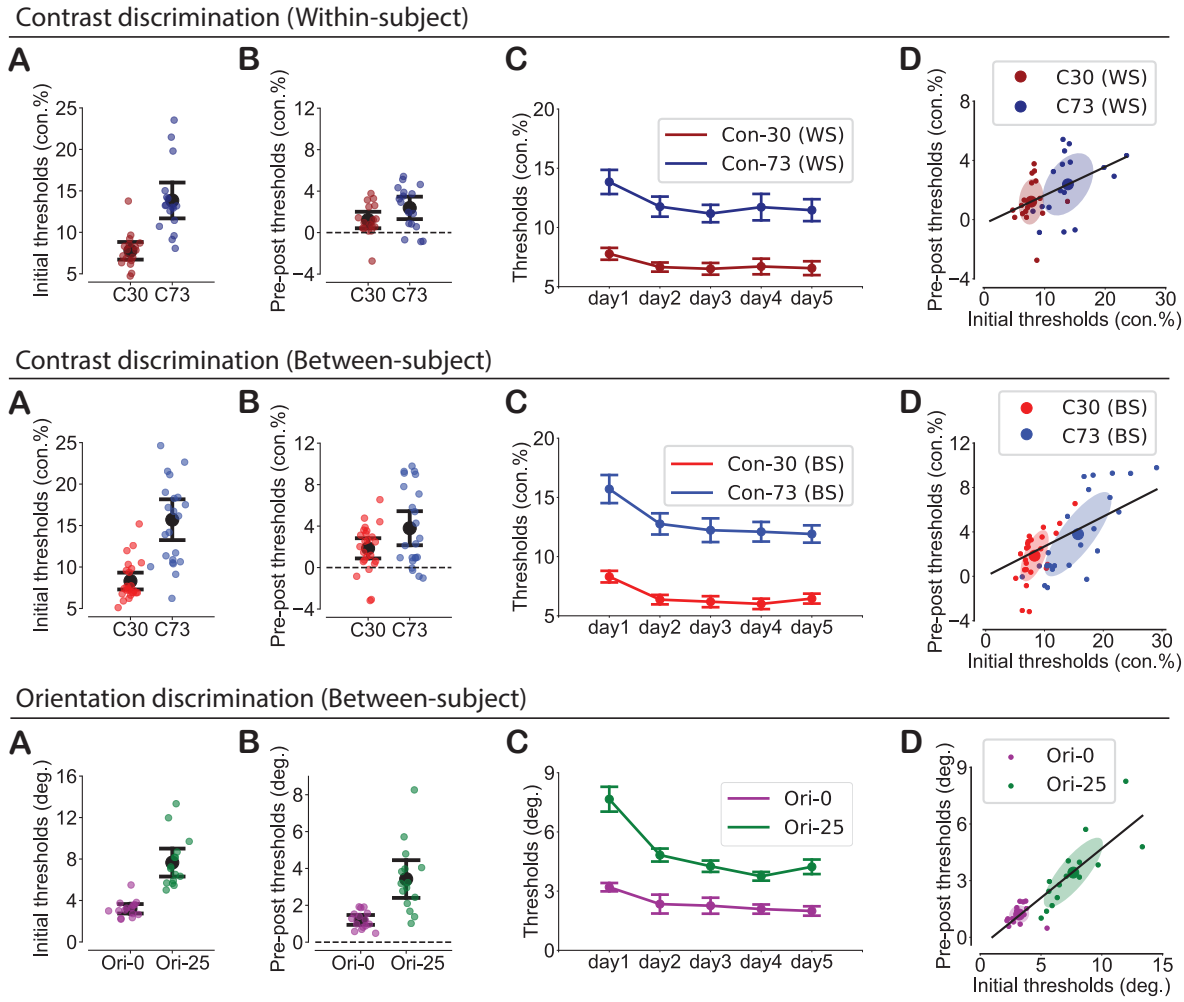


Figure 2.3: Initial discrimination thresholds and the amount learning. **Top panel**: contrast discrimination task, within-subject design. **Middle panel**: contrast discrimination task, between-subject design. **Bottom panel**: orientation discrimination task, between-subject design. In the contrast experiments red color denotes low (con. 30%) and blue color denotes high reference value conditions (con. 73%). In the orientation experiments purple color denotes low (ori. 0°) and green color denotes high reference value conditions (ori. 25°). In all panels: **(A)** Initial discrimination thresholds and **(B)** the amount of absolute learning at the two measured reference values. Error bars represent 95% confidence intervals of the mean. **(C)** Learning curves for the 5-day training protocol for the two measured reference values. Error bars show one SEM. **(D)** Learning as a function of initial discrimination thresholds. Error ellipses show one standard deviation, and black lines show linear regression lines fitted to the points from both conditions. Adapted with permission from J. Fiser and G. Lengyel.

2.2.3.2 Initial performance & learning - within-subject design

In the first contrast discrimination experiment using within-subject design, we tested within each observer directly whether the amount of observers' learning was proportional to their initial thresholds. The proportionality rule states:

$$\frac{IT_{@LowRef}}{IT_{@HighRef}} = \frac{PL_{@LowRef}}{PL_{@HighRef}} \tag{2.1}$$

where $@LowRef$ and $@HighRef$ refer to initial thresholds (IT) or perceptual learning (PL) assessed at the low (con.30% and ori. 0°) or high (con.73% and ori. 25°) reference values. These low and high reference values determined the low and high stimulus base-intensities in our experiments by modulating observers' initial thresholds according to their own perceptual scaling function.

Following Eq. 1, we derived the predicted amount of learning in the low reference value condition ($PL_{@LowRef}$) by multiplying the left side of Eq. 2.1 with the amount of learning in the high-reference-value condition,

$$\frac{IT_{@LowRef}}{IT_{@HighRef}} PL_{@HighRef} = PL_{@LowRef} \tag{2.2}$$

For each participant, we computed the predicted amount of learning (left side of Eq. 2.2) at the higher reference value (high base-intensity) and compared it to the absolute amount of learning (right side of Eq. 2.2, PRE – POST thresholds) at the low reference value (low base-intensity) within the same observer. If the proportional relationship between the initial thresholds and the amount of learning holds, we expect no difference between the predicted and the absolute learning scores. Indeed, we found no difference between the two learning scores (Fig. 2.4, top panel A) confirming the proportional relationship between initial thresholds and learning

(t_{16}=0.216, P=0.832, d=0.049, Bayes Factor favoring no difference=10.6). The error bar in Fig. 2.4, top panel B indicate that most of the observers (13/17) deviated less than 1% contrast from the exact proportionality rule as the observers' amount of learning at the two reference values (base-intensities) was almost exactly proportional to their initial threshold levels. This suggests that the individual perceptual scaling functions dominated quite robustly the origin of the proportionality relationship between learning and initial thresholds. The Bayes Factor indicates directly that the "no difference between the learning scores" hypothesis is 10.6 times more probable than "the existence of a difference between the learning scores" (see Methods, Statistical analyses, comparing group means). Therefore, we found strong evidence for the proportional relationship in the data of Exp. 1, and we linked this relationship directly to observers' perceptual scaling functions.

2.2.3.3 Initial performance & learning - between-subject design

A recurring danger with a within-subject design is the possible confounding effect of cross-training between the conditions, which would allow an alternative explanation to our results in Exp. 1. This calls for an independent confirmation of our findings about proportionality by using a between-subject design. Unfortunately, due to the between-subject design of Experiments 2 & 3 it is not possible to test directly the proportional relationship between learning and the initial thresholds within subjects because separate groups of observers were trained at the two reference values. However, since the initial thresholds were assessed at both reference values in each group, one could use Eq. 2.1 to calculate the *predicted amount of learning* for the untrained reference value condition for each participant in the same way as in the previous section in Experiment 1. The only difference is that when comparing the predicted learning in the untrained reference value condition to the absolute learning in the trained reference value condition, one

needs to use between-subject comparison. To perform this test, first, we computed the predicted amount of learning in the group trained with the high-reference-values (con. 73% and ori. 25°) using Eq. 2.2. by simply multiplying the absolute learning scores of the participants at the high-reference-values with the ratio of their initial thresholds at the two reference values ($\frac{IT_{Con30}}{IT_{Con73}}$ in Exp. 2, and $\frac{IT_{Ori0}}{IT_{Ori25}}$ in Exp. 3). We compared these predicted learning scores to the absolute learning scores of the observers in the low-reference-value conditions (con. 30% and ori. 0°) and found no difference between the two groups' scores (Exp. 2, contrast discrimination task: t_{45}=0.314, P=0.755, d=0.094, Bayes Factor favoring no difference = 7.5; Exp. 3, orientation discrimination task: t_{28}=0.596, P=0.556, d=0.225, Bayes Factor favoring no difference = 4.6, Fig. 2.4, subpanel A in all panels).

Second, we computed the predicted amount of learning in the group trained with the low-reference-values (con. 30% and ori. 0°) derived from Eq. 2.1 by solving it for $PL_{@HighRef}$,

$$PL_{@HighRef} = PL_{@LowRef} \,/\, \frac{IT_{@LowRef}}{IT_{@HighRef}} \tag{2.3}$$

Using Eq. 2.3, we divided the absolute learning scores of the participants at the low-reference-values with the ratio of their initial thresholds at the two reference values ($\frac{IT_{Con30}}{IT_{Con73}}$ in Exp. 2, and $\frac{IT_{Ori0}}{IT_{Ori25}}$ in Exp. 3). When comparing these predicted learning scores to the absolute learning scores of the observers in the high-reference-value conditions (con. 73% and ori. 25°), we found again no difference between the two groups' scores (Exp. 2, contrast discrimination task: t_{45}=0.689, P=0.494, d=0.206, Bayes Factor favoring no difference=5.7; Exp. 3, orientation discrimination task: t_{28}=1.091, P=0.284, d=0.412, Bayes Factor favoring no difference=2.9, Fig. 2.4, subpanel B in all panels).

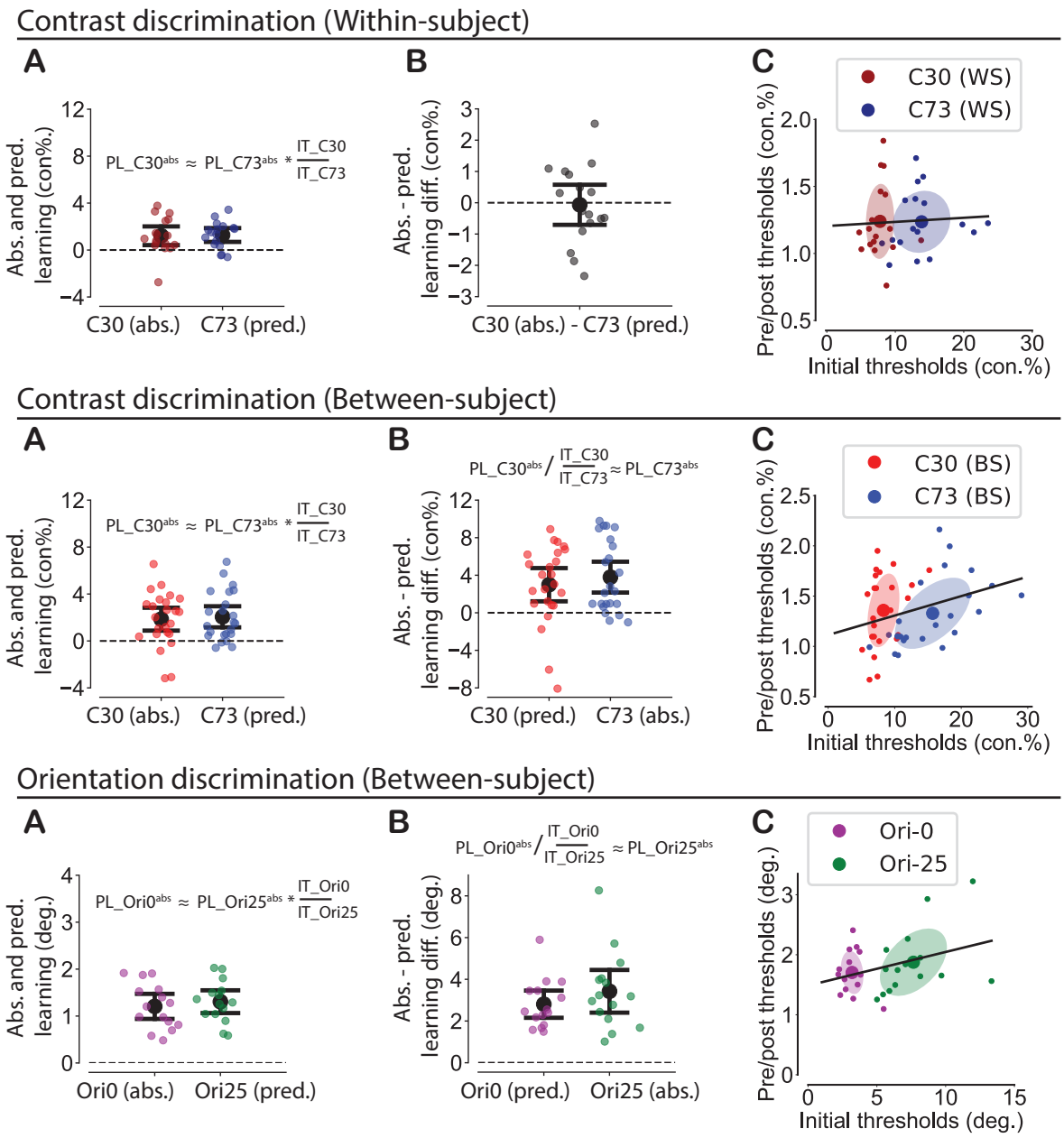


Figure 2.4: The relationship between initial discrimination thresholds and the amount of learning primarily reflects the observers' scaling function. **Top panel**: contrast discrimination task, within-subject design. **Middle panel**: contrast discrimination task, between-subject design. **Bottom panel**: orientation discrimination task, between-subject design. In all panels: **(A)** Comparing the absolute learning in the low-reference-value condition to the predicted learning in the high-reference-value condition. **(B)** Top panel: The difference between the absolute and the predicted amounts of learning at the low and high reference values across subjects. **(B)** Middle & Bottom panels: Comparing the predicted learning in the low-reference-value condition to the absolute learning in the high-reference-value condition. **(A & B)** Error bars represent 95% confidence intervals of the mean, and the equations above the error bars relate absolute to predicted learning in the different conditions derived from Eq. 1 capturing the proportional relationship between initial thresholds and learning. **(C)** Relative learning defined by the ratio of initial discrimination and the post-training thresholds as a function of the initial threshold levels. Error ellipses show one standard deviation, black lines indicate linear regression lines fitted to the points from both conditions. Adapted with permission from J. Fiser and G. Lengyel.

In the Appendix A.2, we provide further explanation as to why these between-subject comparison results support our claim that the amount of learning in these perceptual learning tasks was modulated only by the participants' perceptual scaling function without any additional processes.

2.2.3.4 Individual differences in initial thresholds & learning

We analyzed the individual differences within conditions and investigated how much of the inter-subject variability in learning could be explained by the initial discrimination threshold levels of the observers assuming a proportional relationship between initial performance and learning.

The individual differences in initial performance levels could explain a large part of the variability in learning in all experiments (variance explained in Exp. 1 was 20%, in Exp. 2 was 55%, and in Exp. 3 was 74%, Fig. 2.3, subpanel D in all panels). To test whether the relationship between initial thresholds and the amount of learning was proportional, we computed the relative learning scores of the observers as the ratio of the initial and the post training discrimination thresholds ($\frac{\text{initial threshold}}{\text{final threshold}}$). If the relationship between the amount of learning and the initial discrimination threshold levels is strictly proportional the relative learning scores should be the same at different initial threshold levels. Consequently, there should be no correlation between the relative learning scores and the initial threshold levels. Specifically, $PRE - POST(\text{learning}) = c\,PRE$ with a constant c. Solving this equation for relative learning yields $\frac{PRE}{POST} = \frac{1}{1-c}$, which is a constant again. Following this analysis, in Exp. 1 we found that the positive correlation between learning and the initial thresholds completely disappeared when we used relative learning instead of the absolute learning scores. This suggests that the observers' learning was strictly proportional to their initial discrimination thresholds (Fig. 2.4,

top panel, C, and Table 1). In contrast to Exp. 1, in Exp. 2 & 3 a significant positive relationship between the relative learning and the initial thresholds remained suggesting that the amount of learning in these experiments was not strictly proportional to the initial threshold levels at the inter-subject variability level (Fig. 2.4, middle & bottom panels, subpanel C, and Table 1). On the one hand, this suggests that the relationship between learning and initial performance does not solely reflect the observers' perceptual scaling function from physical to perceptual magnitudes, but there are additional unknown factors strengthening that relationship beyond proportionality. Inter-subject variability is known to reflect arousal level, attention, and motivation (Fahle & Henke-Fahle, 1996; Weiss et al., 1993), each of which can influence the initial discrimination thresholds and can also be modulated by the training causing a positive relationship between learning and initial performance. On the other hand, the correlations were much smaller between the initial discrimination thresholds and the relative learning than between the initial discrimination thresholds and the absolute learning. In Exp. 2, the correlations were, R=0.74 with absolute learning and R=0.31 with relative learning, with a significant difference between them (Z=2.911, P=0.004). In Exp. 3 the same correlations were R=0.86 with absolute learning and R=0.34 with relative learning with an even more significant difference between the two (Z=3.625, P<0.001). This means that even when looking at inter-subject variability, the relationship between learning and initial performance mainly reflects the effect of the perceptual scaling function of the observers. When the influence of the perceptual mapping is factored out by using the relative learning, most of the positive relationship disappears and the explained variance drastically decreases (approximately from 70% to 10%, see the exact correlation coefficients above, and in Table 2.5).

To sum up our findings, the amount of learning was proportional to the initial threshold levels reflecting the effect of the observers' perceptual scaling function linking physical and

perceptual magnitudes. This effect fully captured the observed relationship found in the within- and between-subject analyses when comparing the group means of the conditions with different stimulus base-intensities, and it also explained most of the individual variation between the participants within conditions.

Correlations between initial thresholds and learning (pre-post thresholds)			
Experiment	Correlation coefficient	95% Confidence interval	p value
Exp 1. contrast within-subject design (Fig. 3D, top panel)	r = 0.45	CI95 = 0.12 - 0.69	p = 0.007
Exp 2. contrast between-subject design (Fig. 3D, middle panel)	r = 0.74	CI95 = 0.57 - 0.85	p < 0.001
Exp 3. orientation between-subject design (Fig. 3D, bottom panel)	r = 0.86	CI95 = 0.72 - 0.93	p < 0.001
Correlations between initial thresholds and relative learning (pre/post thresholds)			
Exp 1. contrast within-subject design (Fig. 4C, top panel)	r = 0.05	CI95 = -0.30 - 0.40	p = 0.76
Exp 2. contrast between-subject design (Fig. 4C, middle panel)	r = 0.31	CI95 = 0.01 - 0.55	p = 0.035
Exp 3. orientation between-subject design (Fig. 4C, bottom panel)	r = 0.34	CI95 = -0.03 - 0.63	p = 0.061

Figure 2.5: Analyzing inter-subject variability with correlation. con-ws stands for contrast discrimination with within-subject design. con-bs stands for contrast discrimination with between-subject design. ori-bs stands for orientation discrimination with between-subject design. Adapted with permission from J. Fiser and G. Lengyel.

2.2.3.5 The relationship between learning & generalization

The second goal of our study was to investigate whether the extent of generalization is proportional to the amount of learning in our paradigms. To this end, we analyzed inter-subject variability and found positive correlations between the amount of learning and the extent of generalization in all of the experiments (see Table A.6, and Fig. A.5).

Since the inter-subject variability was much smaller when the reference orientation was at the cardinal orientation compared to the variability at 25°, the above correlational analysis could be misleading due to the large differences in the variances of the learning and generalization scores (see Fig. A.5G & H). Therefore, we included two additional groups of observers in the orientation discrimination experiment. The observers underwent the same experimental

protocol except that one group practiced with reference orientation 15° and the generalization of learning was assessed at 45°, while the other group practiced with 45° reference value and the transfer of learning was measured at 15° (see Appendix A.3). In these groups, the inter-subject variability was similar at both reference orientations (45° and 15°) and it was also large enough to study correlation between generalization and learning (Fig. A.5I & J).

Beside the positive relationship between learning and generalization we also found positive correlations between the initial threshold levels and the amount of generalization in all experiments (see Table A.6, and Fig. A.5). Since the measurement of generalization is also based on the estimation of the discrimination thresholds, observers' perceptual scaling function from physical to perceived magnitudes should influence the amount of generalization in the same way as it influences the amount of learning (see Fig. 2.1 for explanation). This would automatically imply a positive relationship between initial threshold levels and the extent of generalization. However, we were interested in the relationship between learning and generalization without the obvious common influence of the initial discrimination thresholds. Therefore, we computed the partial correlations between learning and generalization while controlling for the initial threshold levels. Despite factoring out the effect of the initial thresholds, we found positive correlations in all experiments (Table 2.6, and see section 2.2.2, Method, Analysis for more information about partial correlation). These findings validate our results suggesting a positive relationship between the amount of learning and generalization in all experiments and confirms that the observed correlations were not due to the self-evident modulating effect of the initial discrimination thresholds.

Partial correlations between learning and generalization while controling for initial thresholds				Partial correlations between learning and relative generalization while controling for initial thresholds		
Experiment	Correlation coefficient	95% Confidence interval	p value	Correlation coefficient	95% Confidence interval	p value
Exp 1. transfer to con. 47% from con. 30%	r = 0.65	CI95 = 0.23 - 0.87	p = 0.004	r = -0.24	CI95 = -0.56 - 0.29	p = 0.356
Exp 1. transfer to con. 47% from con. 73%	r = 0.46	CI95 = 0.05 - 0.77	p = 0.091	r = 0.05	CI95 = -0.45 - 0.53	p = 0848
Exp 2. transfer to con. 73% from con. 30%	r = 0.46	CI95 = 0.00 - 0.76	p = 0.024	r = -0.14	CI95 = -0.57 - 0.35	p = 0.521
Exp 2. transfer to con. 30% from con. 73%	r = 0.20	CI95 = -0.30 - 0.61	p = 0.366	r = -0.25	CI95 = -0.64 - 0.25	p = 0.259
Exp 2. transfer to con. 47% from con. 30%	r = 0.73	CI95 = 0.40 - 0.89	p < 0.001	r = -0.20	CI95 = -0.61 - 0.29	p = 0.416
Exp 2. transfer to con. 47% from con. 73%	r = 0.55	CI95 = 0.11 - 0.81	p = 0.015	r = -0.45	CI95 = -0.76 - 0.02	p = 0.053
Exp 3. transfer to ori. 25˚ from ori.0˚	r = 0.59	CI95 = -0.03 - 0.88	p = 0.020	r = 0.42	CI95 = -0.26 - 0.82	p = 0.116
Exp 3. transfer to ori.0˚ from ori.25˚	r = 0.14	CI95 = -0.52 - 0.62	p = 0.622	r = 0.33	CI95 = -0.36 - 0.78	p = 0.233
Exp 3. transfer to ori. 15˚ from ori.45˚	r = 0.25	CI95 = -0.43 - 0.75	p = 0.454	r = -0.05	CI95 = -0.64 - 0.58	p = 0.891
Exp 3. transfer to ori. 45˚ from ori.15˚	r = 0.89	CI95 = 0.62 - 0.97	p < 0.001	r = -0.10	CI95 = -0.54 - 0.67	p = 0.761

Figure 2.6: Top: Partial correlations between learning and absolute generalization. **Bottom**: Partial correlations between learning and relative generalization computed as generalization divided by the amount of learning. Notes: Transfer to con.47% from con. 30% denotes the condition in which training were at reference value con.30% and generalization was measured at con.47%. The notations for the other conditions follow the same logic. Adapted with permission from J. Fiser and G. Lengyel.

We also computed the relative generalization for each observer by taking the ratio of the amount of learning and the extent of generalization. If generalization is proportional to the amount of learning, this relative generalization should be constant at different amounts of learning because the proportionality relationship claims that $generalization = c learning$; thus $\frac{generalization}{learning} = c$, where c is a constant. Indeed, using relative generalization, the positive correlations we found with the absolute generalization scores vanished and became statistically indistinguishable from zero (Table 2.6).

One potential caveat of this analysis is related to the fact that generalization was assessed by comparing the performance in the untrained conditions at pre- and posttest. If there were no learning between day two (first day of practice) and five (posttest) it would raise the possibility that the measured generalization scores mainly reflect the influence of the pretest which cannot be considered as true generalization because it is identical for the trained and untrained conditions. Indeed, looking at the learning curves in Fig. 2.3 subpanels C, it is evident that

most learning took place from Day 1 one to Day 2 in most experiments. However, our analyses revealed that there was still a significant improvement in most of the conditions after the second day of practice (see Appendix A.4, Fig. A.4). This means that the measurement of generalization used in the present study truly assesses generalization, even if it most probably overestimates somewhat its magnitude. Based on these measurements, our data support the claim that the extent of the generalization in our experiments was proportional to the observers' learning.

2.2.4 General discussion

In three experiments, we investigated (1) how initial performance, as quantified by discrimination threshold at pretest, and overall learning performance were related, and (2) how learning performance and ability to generalize were linked in customarily used perceptual learning tasks. Our goal was to identify general rules that apply to a wide range of conditions during perceptual learning. First, we confirmed the Weber-like law relationship between the initial threshold levels and the amount of learning reported by Astle et al. (2013) and showed that it essentially reflects the perceptual scaling function of the observers without any evidence of additional learning-related processes. Moreover, we found that this proportionality relationship explained not only group mean results but also most of the individual variation across participants. Second, we found that the extent of generalization was proportional to the amount of observers' learning. In the following, we relate our results to the earlier literature and reflect on the implications of the present findings.

2.2.4.1 Initial performance & learning

First, we discuss the comparison of the low- and high-reference-value (i.e. stimulus base-intensity) conditions and how these results relate to the earlier findings of Astle et al. (2013). Second, we consider the results coming from the inter-subject variability analysis and discuss its relation to previous studies (Aberg & Herzog, 2009; Astle et al., 2013; Fahle, 1997; Fahle & Henke-Fahle, 1996).

The results of Astle et al. (2013) and the current experiments are in agreement: they both show proportionally more learning in the conditions with higher initial thresholds compared to conditions with lower initial thresholds. Astle and his colleagues (2013) used monocular, single-interval Vernier acuity task with a 10-day long training protocol and they modulated the initial discrimination threshold levels by changing the eccentricity of the stimuli in a within-subject design. We applied binocular, two-interval contrast and orientation discrimination tasks with a 5-day long training protocol and the initial discrimination threshold levels were modulated by changing the reference contrast and orientation values in within- and between-subject designs. Astle et al. (2013) also showed that the modulation of the initial performance level with crowding or with changing the size of the stimulus elicits the same effect on the amount of learning. The present study used different reference values to modulate initial performance, which again showed a very similar effect on the amount of perceptual learning. Regardless of these differences, in both studies across six experiments, the amount of learning was proportional to the initial thresholds. Since these two studies found consistent results across three different paradigms, under two different training protocols, by using different factors for modulating initial performance levels, together they point towards a general rule in perceptual learning that can predict the amount of learning from the initial discrimination threshold levels. Specifically, regardless of what mechanism constrains the visual discrimination thresholds the amount of

learning will be proportional to the initial thresholds (Astle et al., 2013).

Regarding the origin of this proportionality rule, Astle and his colleagues' (2013) interpretation is quite different from ours. They proposed that the same cortical factors that put a limit on visual perception determining the discrimination thresholds constrain the amount of learning resulting in a Weber-like law during perceptual learning. We found that there was no extra constraint by any cortical factor that modulated learning in addition to the known perceptual processes. Rather, when perceptual scaling was considered at the individual level, the Weber-like law between initial thresholds and learning naturally emerged without any further assumptions. This implies that, after the transformation of the input from the stimulus space to perceptual space takes place, the same amount of perceptual learning occurs at all stimulus intensity levels for all lower level visual attributes. Furthermore, the proportional relationship between initial thresholds and learning also implies that there was no change in the shape of the observer's perceptual scaling function due to the training, only the resolution got higher at the practiced stimulus intensities (i.e. the perceptual discrimination threshold decreased).

In principle, the proportional relationship between initial threshold and amount of learning could also be explained as a result of a particular combination of change in the shape of the perceptual scaling function and/or additional learning effects beyond the simple perceptual scaling that we suggest here. However, based on parsimony, we find such a complex explanation unlikely.

Considering inter-subject variability, the amount of learning in our first experiment using a within-subject design was strictly proportional to the individual initial threshold levels in accordance with the results of Astle et al. (2013). However, in our other two experiments using a between-subject design, the amounts of learning increased more rapidly as a function of the initial threshold levels surpassing proportionality in line with the previous findings of (Aberg &

Herzog, 2009; Fahle & Henke-Fahle, 1996). Exploring this discrepancy, we found that most of the variance in the relationship between initial discrimination threshold levels and learning was captured by the proportionality rule in all of our experiments. Therefore, while other (unknown) factors could also influence the relationship between initial threshold and learning, those represent only secondary effects. We attribute the origins of those secondary, unknown factors to arousal level, attention, and to motivation (Fahle & Henke-Fahle, 1996; Weiss et al., 1993), which can influence the initial discrimination thresholds and can also change due to practice, hence causing a deviation from the strict proportionality rule. Thus, inter-subject variability can also be well explained by the proportional relationship between initial thresholds and learning.

2.2.4.2 Learning & generalization

Considering the link between the amount of learning and the extent of generalization, our results suggest that more learning predicts proportionally more generalization in the standard perceptual learning paradigms with 5-day training. This proportionality relationship was supported by (1) the significant positive partial correlations between the amount of learning and the extent of absolute generalization while controlling for different initial threshold levels, and (2) by the NON-significant partial correlations between the amount of learning and the extent of relative generalization (while controlling for initial threshold levels).

We can reconcile the contradiction between this conclusion and earlier reports showing more learning but less generalization after longer training (Hussain et al., 2012) by considering the two components of learning mentioned in the introduction: the specific characteristics of the training data and overfitting. Depending on whether or not the training data represents the space of the task well, acquiring more knowledge about this training set can help with generalization or hinder it. However, adjusting the internal model of the learner excessively to a

training set regardless of how well it represents the space of the task (i.e. overfitting the data) will necessarily lead to less generalization. The interplay between these two components in the specific setup of (Hussain et al., 2012; Jeter et al., 2010) led to a lack of generalization. This effect might have been due to the increased training length applied in the tasks of Jeter and colleagues' study (2010) since encountering more training trials from the same kind increases the chances of overfitting (observers adjust their internal model more tightly according to the frequently observed trials). In contrast, the training length (in number of trials) in our experiments was fixed at about half of that used in the longest session of Jeter et al. (2010) implying less overfitting and more generalization. Therefore, our training protocol might have created a condition that did not overrepresent particular aspects of the space of the task as much as in previous studies (Hussain et al., 2012; Jeter et al., 2010) leading to the observation that the more observers learned the more they generalized. Since a number of factors related to the task and the stimuli can influence when overfitting begins, the nature of specificity or transfer of learning might not be related to the amount of learning directly, but rather to the balance between the extent of learning, stimulus variability, and the given task with its specific features (Hussain et al., 2012; Jeter et al., 2010). Clearly, this hypothesis of ours suggesting that it is the stronger overfitting and not the larger amount of learning per se that is responsible for specificity in standard perceptual learning tasks remains to be tested in future studies.

2.3 Conclusion

The present study investigated two simple, but general rules that can predict performance in perceptual learning paradigms. First, we confirmed that initial performance and learning are related through a Weber-like relationship regardless of the learning task and showed that this link is a direct consequence of the observers' perceptual scaling function relating physical intensities

to perceived magnitudes. Second, we found that the more people learn under the typical 5-day training protocol the more they generalize. This implies that enhanced specificity reported in some previous studies were not an inherent consequence of the paradigm of perceptual learning with repetitive training but rather of overfitting the training set which is determined by a number of additional factors of the experimental design.

These findings suggest that the patterns in the task and the stimuli are essential to capture learning and generalization in every PL task and there is no general rule that would determine PL performance across the board. This supports the main proposal this thesis (see section 1.4) stating that computational models of PL should incorporate the top-down influence of learning task structures and regularities in the stimuli and that the hierarchical Bayesian approach can inherently combine those top-down influences with the bottom-up sensory observations providing an ideal computational framework for PL.

In the next chapter I will investigate the generalization effects in classical statistical learning paradigms and show that when participants learn the structure embedded in the stimuli, they create abstract, object-like representations that serve as units for allocating attention and that allow generalization to a completely untrained modality.

Chapter 3

Object-based Attention & Across-modality Generalization in SL

3.1 Summary

Although objects are the fundamental units of our representation interpreting the environment around us, it is still not clear how we handle and organize the incoming sensory information to form object representations. The concept of objects is usually defined by a consistent set of sensory properties and physical affordances. Most accounts assume that visual or haptic boundaries are crucial in creating object representations. But how do boundaries emerge and why is the cognitive process of perceptual organization sensitive to those luminance contours? This chapter investigates the hypothesis that boundaries are not essential a priori for the emergence of object concepts, but simply reflect a more fundamental principle: consistent visual or haptic statistical properties.

First, we investigated object-based processing in a classical spatial statistical learning (SL) paradigm. By utilizing previously well-documented advantages of within-object over across-

object information processing, we tested whether learning involuntarily consistent visual statistical properties of stimuli, that are free of any traditional segmentation cues, might be sufficient to create object-like behavioral effects. We combined a classical spatial SL paradigm with measuring efficiency in a novel 3-AFC search task (Study 2, Experiment 1) and with measuring the attentional effect of cueing in the classical object-based attention paradigm (Study 2, Experiment 2). We found that statistically defined and implicitly learned visual chunks bias observers' behavior in subsequent search tasks the same way as objects defined by visual boundaries do.

Second, we investigated generalization between the visual and haptic domains in a novel visuo-haptic statistical learning paradigm. We familiarised participants with objects defined solely by across-scene statistics provided either visually (Study 3, Experiment 1) or through physical interactions (Study 3, Experiment 2). We then tested them on both a visual familiarity and a haptic pulling task, thus measuring both within-modality learning and across-modality generalisation. Participants showed strong within-modality learning and strong 'zero-shot' across-modality generalisation which were highly correlated.

The results of the two studies demonstrate that humans can segment scenes into abstract chunks, without any explicit boundary cues, using purely statistical information. Furthermore, the learnt chunks elicited similar behavior effects as true objects with explicit visual boundaries suggesting that learning consistent statistical contingencies based on the sensory input contributes to the emergence of object representations. We argue that the generalization effects observed in these studies are in line with the hierarchical Bayesian framework (see section 1.4 in the introductory chapter) supporting a probabilistic chunking mechanism. The two studies presented in this chapter were published in Lengyel et al. (2021) (Study 2) and in Lengyel et al. (2019) (Study 3).

3.2 Study 2 - Object-based attention

3.2.1 Introduction

Instead of perceiving the environment as continuous parallel streams of different information flows, our brain organizes the incoming sensory information into meaningful, distinctive units, called objects, and events determined by causal relationships between these objects. Thus, forming internal representations of objects is fundamental to our perception, and understanding this process is an important step toward developing abstract concepts in the human brain. Yet, it is still unknown what object representations are and how they emerge based on processing and organizing the incoming sensory information.

There is an intensive debate in the field about the cues that are necessary and/or sufficient to form the percept of a visual object dominated by earlier results in object cognition, which demonstrated that stable boundaries defined by luminance contours are one of the strongest criteria for visual "objectness" (Kellman & Spelke, 1983; Palmer & Rock, 1994; Spelke, 1990). Indeed, traditional definition of object representations starts with segmenting the objects from the rest of the input based on boundary information (Heydt et al., 1984; Marr, 1982; Peterson, 1994; Riesenhuber & Poggio, 1999; Zhou et al., 2000). However, similarly to segmenting individual words within a continuous speech during hearing (Aslin, 2017; Aslin et al., 1998; Saffran et al., 1996), segmenting objects from the background is an unresolved challenge in vision as most natural experiences contain ambiguous information about object boundaries leading to a large number of potentially correct segmentations (Feldman, 2003; Sun & Fisher, 2003). Just as apparent pauses are bad predictors of word endings in speech (Cole, 1980; Lehiste, 1970), visual edges, contrast transitions, and changes in surface textures are notoriously difficult to identify, and tracking them can lead to false object boundaries (Kanizsa, 1979; Kellman &

Shipley, 1991; Palmer, 2002). In real-life situations, relying exclusively on specific low-level perceptual cues (such as edges) in the sensory input has been proven to be insufficient for finding the true objects in the environment (Kellman & Shipley, 1991; Spelke, 1990).

One potential solution to this problem is based on the proposal that it is not edge boundaries that are required for object definitions but instead, they manifest just one (albeit important) example of a more general principle that leads to object representations: consistent statistical properties co-occurring in the input (Lengyel et al., 2019). Such multi-faceted statistical properties might be more ubiquitous, more reliable to detect and, instead of being encoded innately, a large fraction of them can be learned from and tuned by experience similar to how statistical cues help babies to successfully segment speech (Erickson & Thiessen, 2015; Newport, 2016). While this proposal can explain why a wide variety of cues (e.g. disparity [Julesz, 1971; Spelke, 1990] symmetry [Feldman, 2000], or motion [Spelke, 1990]) were found to be sufficient to elicit the percept of an object, it has not been systematically evaluated in the past.

To investigate the emergence of object representations and evaluate the relevance of consistent statistical properties in this process, we used the following rationale. If consistent statistical properties acquired by learning are indeed fundamental in forming object representations, then a set of newly learned arbitrary statistical contingencies, even if they are not connected to traditional cues and even if they are learned implicitly, should manifest the same kind of object-based behavioral-cognitive effects as true objects do. To test this hypothesis, we started with an implicit learning paradigm called visual statistical learning (VSL), which uses a set of artificial shape stimuli to create novel scenes (Fig. 3.1a, VSL - Block 1). Crucially, the only relevant statistical contingencies defining the structure of these scenes are the co-occurrence statistics of the shapes (i.e. the stable shape-pairs in fixed spatial relation forming the scenes) with no link to low-level visual cues (Fiser & Aslin, 2001; Fiser, 2009). Therefore, the low-level contrast

edges, texture transitions or Gestalt structures that can be important in forming classical object boundaries (Feldman, 2003; Geisler et al., 2001; Palmer, 2002) cannot reveal the statistical structure of the chunks in these scenes. Nevertheless, since these chunks are defined by stable statistical contingencies, according to our hypothesis, they qualify as newly learned objects, and therefore, they should induce object-based perceptual effects.

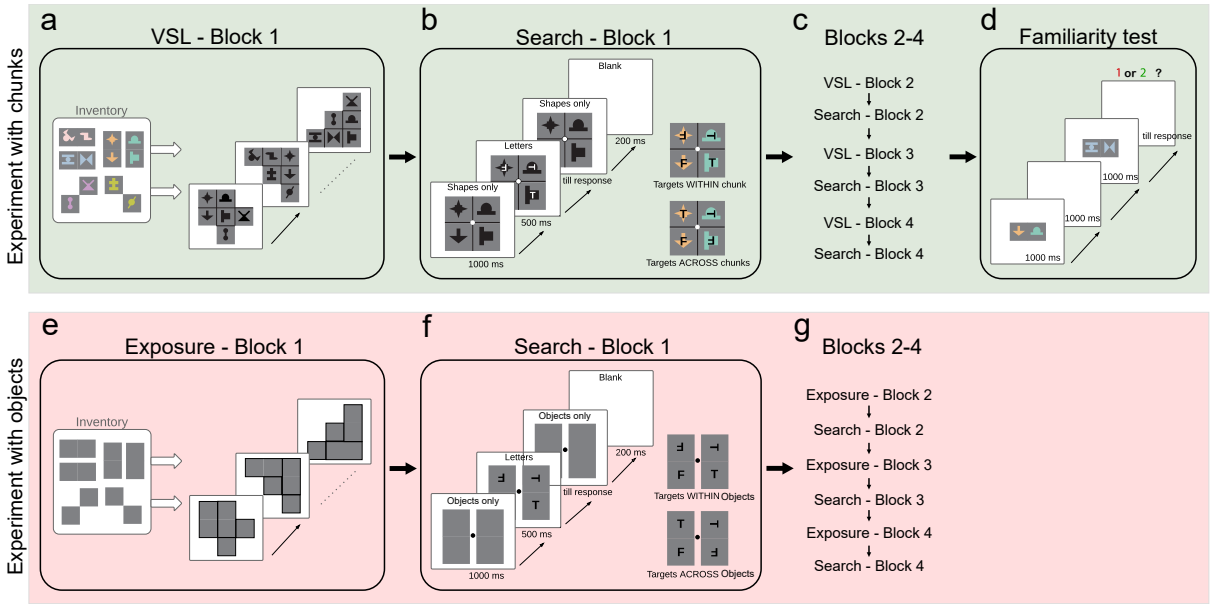


Figure 3.1: The stimuli, the tasks, and the design of Experiment 1. a-d: The design of Experiments 1a and 1b using statistical chunks defined by co-associated abstract shapes. In the Exposure blocks (**a**, VSL-Block 1), true-pairs (Inventory) were used to generate 144 complex scenes for passive viewing. In the Search blocks (**b**, Search - Block 1), observers performed a letter search task with white letters superimposed on the shapes, where the two target letters could be within or across pairs (**b**, Inset, using black letters for visibility). Exposure and Search blocks were presented in an alternating manner (**c**, Blocks 2-4). After the last Search block, a standard VSL Familiarity test was administered to measure the observer's bias to true chunks over random combinations of elements (**c**, Familiarity test). Coloring of the shapes in this figure is only for demonstration purposes, all shapes in the displays were shown in black with no indication of chunk identity. **e-g:** The design of Experiment 1c using objects defined by visual boundary cues. In the Exposure blocks (**e**, Exposure - Block 1), rectangles and squares were used that corresponded to the silhouettes of the pairs in Experiments 1a & 1b. In the Search blocks (**f**, Search - Block 1), observers performed a letter search task with letters appearing in separated rectangles and squares. **f** Insets show trials with within (top) and across (bottom) object setups of targets. The block design of Exp 1c followed that of Experiments 1a & b (**g**, Blocks 2-4). The shapes and the letters are magnified in the figure compared to the actual experimental displays. Adapted with permission from G. Lengyel, M. Nagy, and J. Fiser.

We measured object-related perceptual effects in our scenes with statistically defined ob-

jects in two paradigms. In the first experiment, we design a novel task following previous studies showing that features within an object are detected better than the same features across two objects (Baylis & Driver, 1993; Duncan, 1984; Luck & Vogel, 1997; O'Craven et al., 1999; Vecera et al., 2000). We tested whether observers detected a pair of target letters better when they appeared within a chunk than across two chunks that had been learned implicitly in a preceding VSL session. In the second experiment, we used the well-documented object-based attention (OBA) paradigm (Egly et al., 1994). This paradigm has been used to show that observers responded faster in a cue-based detection task when the target appeared within the object that a preceding cue had indicated compared to when the target appeared on a previously uncued object (Egly et al., 1994; Lee et al., 2012; Moore et al., 1998; Shomstein & Yantis, 2004; Vecera, 1994). We tested whether the same attentional bias would also emerge when instead of objects defined by visual boundaries, the paradigm was applied to newly learned chunks defined by statistical contingencies of abstract shapes. Note that the object-based perceptual effects we measured in these two paradigms were previously attributed exclusively to objects defined by visual boundaries in an explicit manner. Both of our experiments provided clear evidence that recently and implicitly learned statistical chunks without any visual boundary defined by luminance or other traditional cues elicited the same object-based effects as objects with explicit boundaries did.

3.2.2 Experiment 1 - Object-based error rate effect

In Experiment 1, we tested whether the internal representation of the statistical structure developed during a standard VSL paradigm (Fiser & Aslin, 2001) could bias the subsequent visual search task similarly to how objects defined by explicit visual boundary cues would. In alternating blocks of VSL and search trials, observers were exposed to a series of scenes composed

of abstract shapes (Fig. 3.1a-d, in green background). Unbeknownst to the observers, the shape compositions in all the scenes followed a predefined structure based on permanent shape pairs (Fig. 3.1a, VSL - Block 1, Inventory). After each VSL block, observers completed a letter-search task with scenes composed of shapes and letters superimposed on shapes (Fig. 3.1b, Search - Block 1). In each trial, participants had to judge in a three alternative forced choice (3-AFC) task whether they saw (1) two target letters horizontally arranged next to each other, (2) two target letters vertically arranged on top of each other, or (3) just one target letter. If the shape pairs (chunks) that could only be learned from the co-occurrence probabilities of the shapes during VSL blocks behave similarly to objects, then the letter search should be facilitated in this setup by the chunks the same way as it would be by contour-based objects. Indeed, we found that observers detected the targets better when they appeared within a chunk than across chunks both in Experiment 1a and in its replication, in Experiment 1b. These results reflected implicit learning processes and not intentional cognitive strategies, since we excluded from the analysis participants who gained explicit knowledge of the chunks during the experiment (one participant from Experiment 1b, see sections 3.2.2.1 & 3.2.2.1 in the Methods for details). Moreover, when we ran a control experiment, Experiment 1c, which was identical to Experiment 1a & b except that we used objects defined by visual boundaries not chunks (Fig. 3.1e, f, g, in red background), we obtained a behavioral pattern in the visual search task, which was very similar to what we found with statistical chunks.

3.2.2.1 Methods

Participants

Eighty-one university students (53 female, mean age = 21, range = 18-29, 71 right -handed, 49 had normal vision without correction) gave informed consent prior to participation in the

experiment. Thirty participants took part in Experiment 1a, 31 in Experiment 1b (replication) and 20 in Experiment 1c (control). We excluded one participant from Experiment 1b, who explicitly noticed the statistical structure of the pairs (s/he could recall the pairs and the shapes consisting the pairs during the debriefing, see section 3.2.2.1, Methods, Debriefing), since we were interested in the effects of implicit automatic processes and not the consequence of explicit cognitive knowledge. All participants had normal or corrected-to-normal vision. The experimental protocols were approved by the Ethics Committee for Hungarian Psychological Research.

Stimuli

Similarly to previous studies 24,69, the stimuli in the visual statistical learning (VSL) and search blocks in Exp 1a-1b consisted of 12 moderately complex 2D abstract shapes (Fig. 3.1a, VSL - Block 1, Inventory). Unbeknownst to the observers, an Inventory of 6 pairs were constructed from these shapes creating two horizontally, two vertically, and two diagonally oriented pairs. These pairs were the building blocks of the scenes throughout the experiments, as the two elements of a given pair always appeared together in the prespecified spatial configuration defined by the Inventory. Hence, each pair constituted one statistical chunk in our experiment. For each observer, the shapes were randomly reassigned to the pairs in the inventory to eliminate any specific learning effect across subjects due to particular shape combinations.

Tasks and procedure

Observers completed 4-4 (in Experiment 1a and in Experiment 1c) and 2-2 (in Experiment 1b) alternating blocks of VSL and Search trials. Both Experiments 1a-1b were completed with a final Familiarity test and a debriefing, whereas in Experiment 1c such a Familiarity test was omitted as it was not meaningful (Fig. 3.1).

Visual statistical learning paradigm

Observers watched a series of scenes, each constructed from three pairs chosen pseudo randomly from the Inventory (Fig. 3.1a, VSL - Block 1). In each scene, one pair was selected from each of the three types (horizontal, vertical, and diagonal). The three selected pairs could appear in a 3-by-3 grid and their positions were randomized with the constraint that each pair had to be adjacent by side to at least one other pair. This method yielded 144 unique scenes with each pair appearing 72 times during each VSL block. We split the possible scenes into two sets so that each pair appeared in each set 36 times, and presented the two sets alternating: the first set was presented in blocks 1 and 3, while the second set in blocks 2 and 4, all in different randomized order across observers. Each scene was presented for 2 sec with 1 sec pause between scenes. The task of the participants was simply to observe the scenes passively so that they could answer some questions related to their experience afterwards.

Visual search paradigm

After each VSL block, observers had to complete a block of search trials. In these blocks, four shapes were presented in each trial adjacent to each other in a 2-by-2 arrangement (Fig. 3.1b, Search - Block 1). The scenes could contain two true pairs (the two horizontal or two vertical pairs of the Inventory), or one true pair and two individual shapes chosen randomly from the remaining shapes of the two diagonal pairs. The chunks of diagonal pairs were sacrificed in the search task in order to get more possible unique scenes with a 2-by-2 configuration. In each search block, we presented 144 scenes in random order, from which there were 96 unique scenes containing one true pair and two individual shapes (from the two diagonal pairs) and 4 x 12 unique scenes consisting of two true pairs (the horizontal and vertical pairs). All individual shapes were presented 48 times during the search blocks and all horizontal/vertical pairs were presented an equal number of times.

In each search trial, a small white letter appeared in the middle of each black shape, which could be either a T or an F. The task of the observers was to look for the letter Ts among distractor letters (F), and in a 3-AFC task, they had to press 1 on the keyboard if they saw two Ts horizontally arranged next to each other, press 2 if they saw two Ts vertically arranged on top of each other, and press 3 if they saw only one T. The response key mapping (1-beside, 2-top, 3-one target) and the target letter identity (T or F) was counterbalanced across observers. The letters appeared for 500 ms, then they disappeared and only the shapes were visible until the response (Fig. 3.1b, Search - Block 1). Observers were instructed to always keep their eye on the fixation dot in the middle of the scene. When two Ts appeared, they formulated either a horizontal or a vertical pair, and these pairs were randomly distributed the same number of times across the four possible locations in the 2-by-2 configuration. Similarly, when only one T appeared in a trial, its position was randomly and evenly distributed across the four possible locations (top-left, top-right, bottom-left, bottom-right). Each of the three response types (targets on top of each other, targets beside each other, only one target) occurred 48 times randomly distributed across the block.

Familiarity test

After the last search block of Experiments 1a and 1b (replication), observers completed a 2-AFC task typically administered in VSL experiments. In each trial, they saw two pairs of shapes after each other, and they decided which of the two consecutive pairs seemed more familiar to them based on the experiment (Fig. 3.1d, Familiarity test). The two pairs were presented sequentially for 1 sec each with 1 sec pause between them. One of the pairs was a true pair (a horizontal or a vertical pair chosen from the Inventory; Fig. 3.1a, VSL - Block 1, Inventory, top four pairs), while the other random pair was constructed from two shapes arbitrarily chosen from the diagonal pairs (Fig. 3.1a, VSL - Block 1, Inventory, bottom two pairs). Observers performed 8

trials, in which one of the horizontal and one of the vertical true pairs were chosen randomly and tested twice against two randomly paired shapes from the diagonal pairs. The presentation order of the true pair and the random pair was counterbalanced across trials, and the presentation order of the trials was randomized individually for each observer.

Debriefing

VSL is an implicit learning task because observers had no task to perform beyond paying attention to the scenes. However, their knowledge of the statistical structure, that they built during the implicit learning task, could become explicit. Since the Familiarity test does not indicate to what extent the responses were based on implicit or explicit knowledge, we conducted a debriefing at the end of Experiments 1a, 1b, and Experiment 2 (see section 3.2.3.1, Methods, Experiment 2) to identify observers with clear explicit knowledge of the statistical structure. Participants were questioned whether they noticed anything about the shapes during the experiment. If they answered 'yes', they were asked further about what they noticed, and if they said something about pairs of shapes being linked, they were asked to name the shapes in each pair that they remembered. Observers who mentioned noticing consistent pairs during the experiments were considered to be explicit learners who were aware of the hidden statistical structure and, therefore their data was excluded from the analysis.

Control experiment

The control experiment, Experiment 1c was identical to Experiments 1a and 1b with the exception that instead of shape pairs, geometric objects defined by explicit visual boundaries were used as inventory elements (Fig. 3.1e, in red background) and there was no Familiarity test at the end. We used rectangles to represent the true horizontal and vertical pairs, and two squares to represent the two constituent shapes of each diagonal pair. In the Exposure blocks, observers

saw the same number of scenes constructed from the same constituents in the same manner as in the scenes of Experiments 1a-1b, but constructed by rectangles and squares instead of the pairs of shapes. Consequently, the global silhouettes of the composed scenes were also identical across the three experiments. The Search blocks were as similar to those in Experiments 1a-1b as possible. Observers completed the same number and type of trials with the same target locations as in the first two experiments: either two horizontal or two vertical rectangles, similarly to trials with two true pairs in Exps 1a-1b, or one rectangle and two squares, similarly to trials with one true pair and two individual shapes.

Data analysis

In all statistical analyses, we performed the standard two-sided frequentist and the corresponding Bayesian tests and drew our conclusion based on both types of tests combined. In the reported results, the value of the Bayes Factor directly reflected how much more probable the alternative hypothesis was compared to the null hypothesis. For computing the Bayes Factor, we used JZS Bayes factor analysis with a scaling factor of $\sqrt{\frac{1}{2}}$ in the Cauchy prior distribution (Ly et al., 2016; Morey & Rouder, 2011; Rouder et al., 2009).

3.2.2.2 Results

Experiment 1a

In the first search block of Experiment 1a, the statistical chunks significantly modulated the visual search task. Observers committed more errors when the target letters appeared across chunks compared to when the targets appeared within a chunk (t_{29}=4.37, P<0.001, d=0.812, Bayes Factor=186, Fig. 3.2a). After the first block, this effect in the error rate disappeared, none of the error rate differences in search blocks 2-4 differed significantly from zero (ts_{29}<1.91, Ps>0.066, ds<0.354, Bayes Factors<1, Fig. 3.2a). The drop in the chunk-based error rate

effect between the first and the other three search blocks was also significant ($F_{3,87}$=7.417, P<0.001, Bayes Factor=539; post-hoc comparisons of Block 1 vs. Blocks 2-4: ts_{29}>3.04, Ps<0.004, ds>0.565, Bayes Factors>8, Fig. 3.2a). Meanwhile, there was no difference in the measured reaction times between within-chunk (targets appeared within a chunk) and across-chunks (targets appeared across chunks) trials across the four blocks (ts_{29}<|1.50|, Ps>0.145, ds<|0.278|, Bayes Factors<1, Fig. 3.2d).

These results indicate that immediately after the first exposure to the novel structured input (1st VSL block), the implicitly learned chunks influenced the accuracy of the observers in the visual search task in the predicted manner: observers detected the two targets more accurately when the target letters appeared within the same statistically defined chunk compared to when they were distributed across two chunks. To confirm that this effect is indeed linked to the implicit learning of the chunks, we calculated the correlation between the chunk-based error rate difference in Block 1 and the amount of statistical learning measured by the Familiarity test, and found a significant effect (R=0.40, CI_{95}=0.03-0.67, P=0.031, Bayes Factor=3, Fig. 3.2g). This supports the idea that effect in the error rates was a direct consequence of the learned statistical structures during the VSL block.

The overall performance of the observers did not differ significantly from chance in the Familiarity test (t_{29}=1.409, P=0.169, d=0.262, Bayes Factor=0.5, Fig. 3.2g, orange error bar on the x axis) despite the fact that, in total, observers were exposed to twice as many exposure scenes as in the classic experiment of Fiser and Aslin (2001). The most probable explanation of this is that the scenes in our experiment were divided into four shorter exposure blocks interleaved with the search blocks, and thus the interleaved visual search blocks interfered with the performance in the Familiarity test. Importantly, given the substantial variability in learning found during the Familiarity test, this non-significance of the overall magnitude of learning was

irrelevant with respect to the two main results found, namely the differential search behavior of within vs. across learned chunks and the significant correlation between the magnitude of the search difference and statistical learning measured in the Familiarity test.

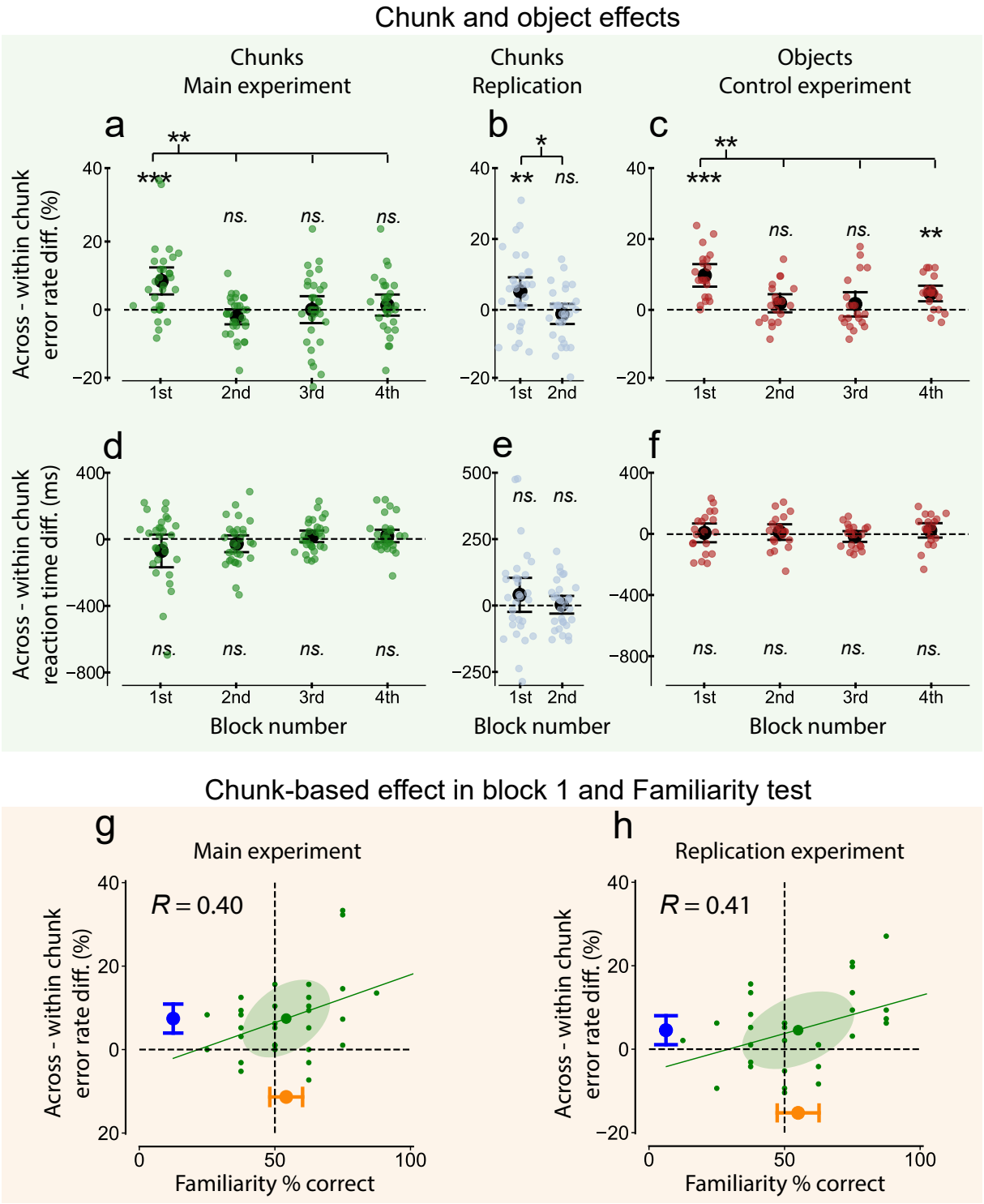


Figure 3.2: Chunk- and object-based error rate effects in Experiment 1. Caption continues on the next page. Adapted with permission from G. Lengyel, M. Nagy, and J. Fiser.

Figure 3.2: (Caption for Fig. 3.2 on the previous page.) **a-f:** Chunk/object-based error rate **(a-c)** and reaction time **(d-f)** effects across Exps 1a, 1b &1c. Mean error rate and median reaction time differences between the across-chunk and within-chunk trials (y axis) in each Search block (x axis) in the main **(a)** and in the replication **(b)**, and in the control **(c)** experiments. Positive values mean fewer errors or faster responses in within-chunk compared to across-chunk trials and error bars show the 95% confidence intervals of the mean. Colored dots represent the mean error rates or median reaction time differences of the observers in a given block. **g, h:** The relationship between performance in the Familiarity test (x axis) and error rate differences of the across-chunk vs. within-chunk trials in the first block (y axis) in the main **(g)** and in the replication **(h)** experiments. Green error ellipses show one standard deviation and green lines represent best-fitting linear regression lines. The error bars show the 95% confidence intervals of the mean performance in the Familiarity test (orange), and of the average chunk-based error rate effect (blue). n=30 in Exp. 1a **(a, d, g)**, n=30 in Exp. 1b **(b, e, h)**, and n=20 in Exp. 1c **(c, f)**. Significant differences from zero in **a-f** are indicated with ns., P>0.05, *, P<0.05, **, P<0.01, ***, P<0.001, two-tailed paired (difference between across and within-chunk trials) t-tests. R values in **g** and **f** indicate Pearson correlation coefficients.

Experiment 1b

To enhance the credibility of our results, we reran Experiment 1a with a different group of observers in Experiment 1b. This time, observers completed only two-two blocks of VSL and search trials since in Experiment 1a, the chunks influenced the performance significantly only in the first search block and it disappeared in the remaining of the blocks (Fig. 3.1a). In the replication Experiment 1b, we obtained exactly the same results as in Experiment 1a (Fig. 3.2b, e): the chunks had a strong effect on error rates in the first search block (t_{29}=2.68, P=0.012, d=0.498, Bayes Factor=4), which disappeared in the second search block (t_{29}=-0.86, P=0.398, d=0.159, Bayes Factor=0.3), the chunk-based effect was also significantly smaller in the second block than in the first search block (t_{29}=2.41, P=0.022, d=0.448, Bayes Factor=2), and there was no effect of the chunks on the reaction times (ts_{29}<1.28, Ps>0.212, ds<0.237, Bayes Factors<0.4).

Using Bayesian statistics, we could combine the data from Experiments 1a and 1b because the first two blocks were identical in those experiments. We computed the probability of the hypothesis that observers made fewer errors in within-chunk trials than in across-chunk trials (following (Morey & Rouder, 2011; Rouder et al., 2009)), and found very strong evidence supporting the existence of the chunk-based effect, with Bayes Factor =2907, indicating that

the existence of a chunk-based effect is 2907 times more probable than assuming no chunk-based effect. Furthermore, the Bayes Factor analysis conducted on the correlations (following (Ly et al., 2016)) indicated that the probability of an existing positive correlation between the chunk-based effect and the performance in the familiarity test was 24 times more probable than assuming no relationship between the two. These results provided further strong evidence that the chunk-based error rate effect was related to the learned statistical structure.

We conducted two additional tests to further strengthen the assessment that implicit learning of the chunks is the driving force behind the error rate effect, and that the significant positive correlation between the chunk-based error rate effect and familiarity is not just due to a generic factor such as attention or across-subject variability in overall performance. First, we computed the partial correlation between the performance in the Familiarity test and the chunk-based error rate effect while controlling for the average performance in the task (measured by individual average error rates), and we found significant positive partial correlations in both experiments (Experiment 1: R=0.39, CI_{95}=0.03-0.67, P=0.028, Bayes Factor=3; Experiment 1b: R=0.38, CI_{95}=0.04-0.66, P=0.015, Bayes Factor=3). This result further corroborates the idea that the underlying cause of the correlation between the performance in the Familiarity test and the chunk-based error rate effect is the implicitly learned statistical structure. Second, we wanted to rule out the possibility that the chunk-based error rate effect emerged solely because observers paid more general attention to the area of the scene with a true-pair structure compared to the area lacking such a structure due to having just two individual shapes (Fig. B.4c). To this end, we repeated the analysis on the subset of trials in the search task, which had two true-pairs and no single elements (i.e. two chunks, see Fig. B.4b). In this case, all positions enjoyed the same advantage from being a part of a chunk, thus the effect had to originate from the targets being within the same chunk. We found the effect of the chunks on the error rate in these trials to have

the same size as in the case of the full set (Experiments 1a and 1b together: t_{59}=3.77, P<0.001, d=0.487, Bayes Factor=64) indicating that the reported chunk-based error rate effect could not be explained by allocating more attention to true-pairs than to individual shapes. In summary, in Experiments 1a and 1b we found convincing evidence that (1) the chunks of the scenes' underlying statistical structure modulated subsequent performance in the visual search task, and (2) this chunk-based error rate effect had a strong positive relationship with the performance in the familiarity test measuring the degree of learning.

Experiment 1c

However, two additional issues had to be clarified for linking these effects to object representations. First, the effect we found diminished after the first search block and second, it is unclear exactly how objects with explicit boundaries would influence the same search in the present 3-AFC paradigm. To address both issues, we ran a control experiment (Experiment 1c), which was identical to Experiments 1a & b in all aspects except that the underlying scene structure was specified by objects defined by visual boundaries instead of chunks defined by abstract shape pairs (Fig. 3.1e, f, g, in red background). Comparing the results of Experiments 1a, 1b, and 1c, we found that objects with explicit visual boundaries elicited a very similar pattern of results to those obtained with statistical chunks (Fig. 3.2c, f and see Fig. B.2). First, objects with visual boundaries influenced the error rates significantly in the first search block, and also significantly more there than in the rest of the search blocks. Specifically, observers made more errors when the targets appeared across compared to within objects in two of the four search blocks (Block 1: t_{19}=6.50, P<0.001, d=1.490, Bayes Factor=6237; Block 2: t_{19}=1.51, P=0.148, d=0.346, Bayes Factor=1; Block 3: t_{19}=0.957, P=0.351, d=0.219, Bayes Factor=0.3; Block 4: t_{19}=4.52, P<0.001, d=1.036, Bayes Factor=130, Fig. 3.2c), but this effect was significantly larger in the first block ($F_{3,57}$=7.709, P<0.001, Bayes Factor=441; comparing Block 1 to Blocks 2-4 post-

hoc: ts_{19}>2.71, Ps<0.014, ds>0.621, Bayes Factors>4, Fig. 3.2c). Second, objects with visual boundaries had no modulatory effect on the reaction times in any of the blocks (ts_{19}<1.10, Ps>0.286, ds<0.252, Bayes Factors<0.4, Fig. 3.2f).

The most parsimonious interpretation of these results is that the reduction of the object/chunk-dependent effect after the first block is due to a floor effect in errors, while the sustained within/across object difference in the later block of Experiment 1c is due to the stronger overall effect obtained by using objects with visual boundaries compared to chunk-based objects. In particular, when observers struggle to learn the task, the effect is the largest both for objects and chunks (1st block), while after having learned the task (blocks 2-4), they make, on average, fewer errors, hence the error difference due to the effect of chunks/objects also decreases. Indeed, we found that in all three experiments, observers made the most errors in the first block and after the first block their performance improved significantly (see section B.1.2, Supplementary material, Experiment 1, Results and Fig. B.4). Thus, while an overall reduction in error difference occurred across the blocks of all three experiments, due to the stronger modulatory effect of objects with visual boundaries, the within/across object difference could still be detected in Block 4 of Experiment 1c using the present paradigm, while it became insignificant in Experiments 1a & b.

More importantly, based on the strong effects we found and the quantitative treatment of the diminishing nature of the effect over time, this set of experiments coherently demonstrated in a novel visual search task that statistical chunks learned in a VSL paradigm elicited very similar behavioral effects to those caused by objects defined by clear visual boundaries.

3.2.3 Experiment 2 - Object-based reaction time effect

If statistical chunks in a VSL paradigm behave as objects defined by explicit visual boundaries, they should also manifest their effect on attention in classical visual cueing paradigms. To test this conjecture and provide further evidence for similar higher-order effects based on objects with visual boundaries and contingency-based novel statistical chunks, we combined the classic object-based attention (OBA) paradigm with the VSL paradigm in our second experiment. Object-based attention (OBA) is a well-documented example of object-related perceptual effects, which is based on reaction time measurements (Egly et al., 1994; Lee et al., 2012; Moore et al., 1998; Shomstein & Yantis, 2004; Vecera, 1994). OBA refers to the phenomenon when an observer's attention is drawn to one part of an object and their attention will automatically include the whole object, not just the part singled out by the cue (Egly et al., 1994; Moore et al., 1998; Vecera, 1994). In the classic demonstration of OBA, observers are asked to identify a target letter among distractor letters in a two alternatives forced choice (2-AFC) task after a partially reliable cue indicates where the target might appear in a scene composed of multiple objects defined by visual boundaries. Observers are faster to identify the target if the cue indicates an incorrect location but the location is within the object tagged by the cue as opposed to the situation, when the target appears not only in an uncued location but also in an uncued object even when the distances of the target from the cue are identical in the two conditions (Fig. 3.3b).

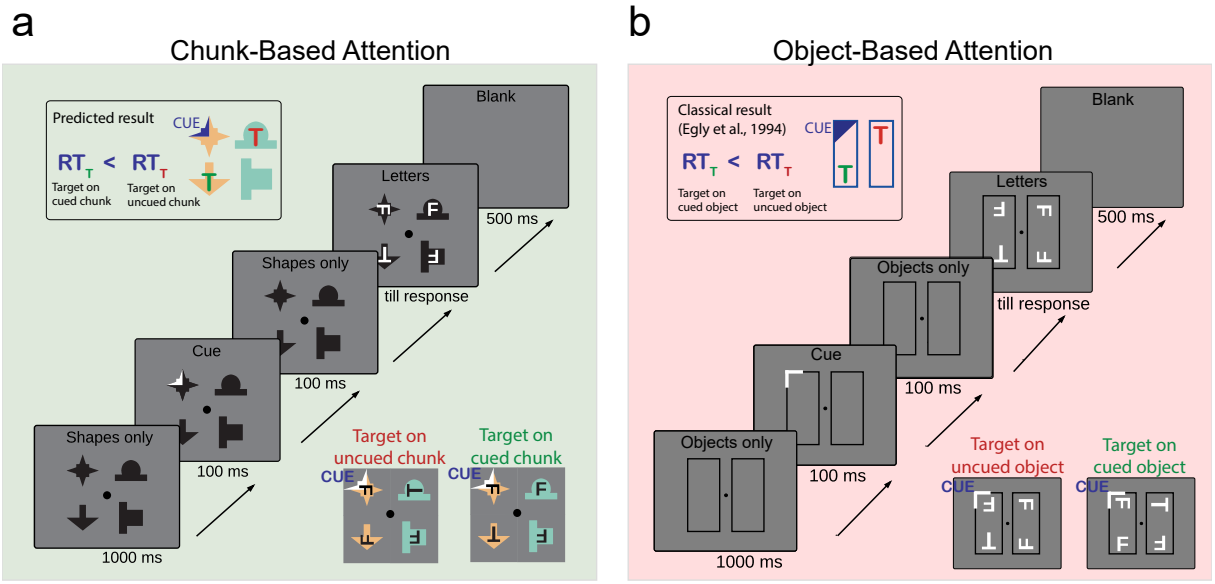


Figure 3.3: The stimuli, the tasks and the trial structures in the Experiment 2. **a** Chunk-based attention paradigm. **b** Object-based attention paradigm. **a, b** Top-left insets display the expected results in the two paradigms (longer RTs when the target appears on the uncued chunk/object vs. cued chunk/object). Bottom-right insets in **a, b** present two examples of trials, in which the target (T) appeared on the cued (green label) and on the uncued (red label) chunks/objects. The design, the visual statistical learning, and the Familiarity test were identical to Exp. 1 (see Fig. 3.1). The shapes and the letters are magnified in the figure compared to the actual experimental displays. Adapted with permission from G. Lengyel, M. Nagy, and J. Fiser.

In order to investigate whether the chunks learned in a VSL task elicit an effect similar to OBA, we followed the same design as in Experiment 1. Observers completed alternating blocks of VSL and OBA trials. In the VSL blocks, similarly to Experiment 1, they were exposed to a series of scenes, which were composed of chunks of shapes (Fig. 3.1a). After each block of VSL, observers completed a set of classical OBA trials (Egly et al., 1994; Shomstein & Yantis, 2004) with one modification: the target and distractor letters appeared superimposed on the shapes of the VSL block, which were arranged in a 2-by-2 configuration (Fig. 3.3a). After finishing all the VSL and OBA blocks, half of the observers completed an additional four blocks of the classic OBA task, but this time using objects defined by explicit visual boundary cues (Fig. 3.3b). This arrangement allowed a direct comparison between chunk- and contour-driven OBA within these observers. Finally, all observers completed a Familiarity test with chunks. We

found that observers identified the targets faster when they appeared at an uncued location which was on a cued chunk compared to when the target was presented on an uncued chunk, replicating the exact same pattern that was found with objects defined by visual boundaries. Furthermore, we found a positive correlation between the OBA and the chunk-based attention (CBA) effect in observers performing both tasks, which suggests overlapping cognitive mechanisms behind the two effects. We also found a strong correlation between the CBA effect and the strength of chunk learning as quantified by the Familiarity test. Similar to Experiment 1, we excluded from these analyses all participants with explicit knowledge about the chunks (5 participants, see sections 3.2.3.1 & 3.2.2.1 in Methods) to assure that our results reflect the consequences of implicit statistical learning and not explicit strategies.

3.2.3.1 Methods

Participants

We estimated the effect size of the original object-based attention (OBA) reported in previous studies and found that, on average, OBA has a small effect size (Cohen's d = 0.22). Since chunk-based attention (CBA) is likely to be even weaker than OBA, we assumed that CBA would yield an effect half as strong as in OBA. Asking for 60% probability to find the CBA, we established that our study required a sample size of 104 observers. We aimed at one hundred observers and managed to recruit 98 university students (68 female, mean age = 21, range = 18-26, 91 right -handed, 60 had normal vision without correction), who gave informed consent prior to participation in the experiments. As in Experiment 1, observers with explicit knowledge of the chunks were excluded (5/98). We excluded three additional observers because they did not finish the experiment, thus data of 90/98 observers were analyzed in this experiment. All observers had normal or corrected-to-normal vision. The experimental protocols were approved

by the Ethics Committee for Hungarian Psychological Research.

Stimuli, tasks, and procedure

The design of the experiment, the stimuli, the VSL blocks, and the familiarity test were identical to Experiment 1 with the exceptions specified below. Observers completed 4-4 alternating blocks of VSL and CBA trials. Due to data acquisition error, 19 observers completed only 3-3 blocks of VSL and CBA trials, but this only reduced the number of trials to 72 from 96 in the experimental conditions, thus their data was used in the analyses. All observers completed a Familiarity test at the end of the final CBA block.

Visual statistical learning paradigm

Based on the assumption that stronger associations lead to larger effects in the CBA task, the number of exposure scenes in the VSL blocks was doubled from 72 to 144 to strengthen the learned associations between the shapes. Set size of 144 was chosen to have a robust learning effect while avoiding an explicit understanding of the input structure. In contrast to the exposure scenes of Experiment 1, the black lines separating the shapes in the scene were completely omitted in order to further decrease lower level visual cues of structure.

Chunk- based attention paradigm

After each VSL block, observers completed a block of CBA trials. In the CBA blocks, four shapes were presented adjacent to each other in a 2-by-2 arrangement without explicit black lines separating them (Fig. 3.3a). The configurations of the different scenes were the same as in Experiment 1: they either contained two horizontal or two vertical true pairs (Fig. 3.1a, VSL - Block 1, Inventory, top four pairs, see also Fig. B.4b) or one true pair and two individual shapes from the diagonal pairs (Fig. 3.1a, VSL - Block 1, Inventory, bottom two pairs, see also Fig. B.4c). Observers were exposed to the same number and mix of scenes as in Exp. 1, and

the scenes were presented in a different random order in each search block.

Following the original OBA method (Egly et al., 1994) in each trial scene, first, only the four shapes appeared for 1000 msec, then one of the shapes was cued for 100 msec. The cue disappeared and only the four shapes were visible for another 100 msec, then one target (R or L) and three distractor letters (F) appeared, one in the middle of each of the four shapes. The letters remained in the center of the shapes until the observer responded. Cueing was provided by coloring a quadrant of the black shape to white (Fig. 3.3a, CBA panel insets). The cue-coloring was designed to draw attention without favoring direction to any location. The observers' 2-AFC task was to press 1 when they saw a letter T, and 2 when they saw an L among the distractor letters F in the given trial. At the beginning of the experiment, they were explicitly instructed to pay attention to the cue as it would correctly predict the location of the target in most, but not all of the trials. Observers were further instructed to continuously fixate at the fixation dot in the middle of the screen. The size of the OBA effect has been found fairly independent of the predictability of the cue in previous studies: similar effect sizes were reported with fully random (Shomstein & Yantis, 2004) and highly predictable cues (Egly et al., 1994). Therefore, the accuracy of the cue in the present study was set to 55%, which was estimated to be sufficient to elicit the OBA effect.

Each CBA block consisted of 144 trial scenes with 80 valid-cue trials (i.e. the cue appeared at the same location as the target), and 64 invalid-cue trials. Of the 64 invalid-cue trials, the target appeared on the cued chunk in 24 trials (Fig. 3.3a, right inset), whereas in the other 24 trials, the target appeared at the same distance from the cued location as in the first 24 trials but in the uncued chunk (Fig. 3.3a, left inset). The remaining 16 invalid-cue trials were used for balancing the frequency of the individual shapes across the block and used only one chunk and two individual shapes in the scene, with the cue appearing in one of the individual shapes. These

trials were not used in the subsequent analysis. The targets and the cues appeared randomly and the same number of times in all four locations of the 2-by-2 layout. In the invalid-cue trials, the target never appeared in the position diagonally opposite to the cued location.

Object-based attention paradigm

49 participants completed 4 blocks of a classic OBA task at the end of the experiment and data of 44/49 observers were analyzed (see exclusion criteria in section 3.2.3.1, Participants). In the OBA blocks, the task was identical to the task in the CBA paradigm, but the target and distractor letters appeared in objects defined by visual boundary cues (i.e. rectangles or squares) instead of the shapes (Fig. 3.3b, OBA panel). We used the boundary-outlined rectangles as objects following previous studies (Egly et al., 1994; Lee et al., 2012; Moore et al., 1998; Shomstein & Yantis, 2004; Vecera, 1994) and augmented those with squares as analogues of the individual shapes constituting the diagonal pairs in the CBA paradigm. Observers completed the same number of trials of the same trial types (either two rectangles -comparable to trials with two chunks- or one rectangle and two squares -comparable to trials with one chunk and two individual shapes) with the same cues, and target locations as in the CBA blocks in a different random order.

Familiarity test

The Familiarity test was identical to the test in Experiment 1 with one modification driven by the goal of increasing the number of trials for a more accurate estimate of learning performance while keeping the appearance frequency of the shapes and pairs balanced. Specifically, we introduced foil pairs and catch-trials in this test in the following manner (see section B.2.1 in the Supplementary material for more information on foil pairs). Observers performed 24 trials in which all true pairs were tested against foil pairs. In each trial the true and foil pairs contained

different shapes. From the 24 trials 16 were normal and 8 were catch-trials. In the catch-trials, observers had to compare two foil pairs. These trials were needed to keep the appearance frequency of the shapes and pairs equal in the Familiarity test. In this way, both the true and the foil pairs appeared four times, and each shape appeared eight times in the test. The presentation order of the trials, and the sequential order of the true and foil pairs in a trial were separately randomized for each subject.

3.2.3.2 Results

Cue validity effect

First, we assessed the standard cue validity effect by measuring how much observers' reaction times and error rates were modulated when the cue indicated the subsequent target position exactly. We found in both the chunk and the object version of the paradigm that observers responded faster (Objects: t_{43}=9.78, P<0.001, d=1.491, Bayes Factor=$5 \cdot 10^9$; Chunks: t_{89}=11.35, P<0.001, d=1.203, Bayes Factor=1016; Fig. 3.4a, left panel), and they made fewer errors (Objects: t_{43}=2.46, P=0.018, d=0.375, Bayes Factor=2; Chunks: t_{89}=4.11, P<0.001, d=0.435, Bayes Factor=217; Fig. 3.4a, right panel) when the target appeared at the cued (valid-cue trials) compared to uncued location (invalid-cue trials). There was no difference between the magnitude of the validity effect in the object vs. the chunk version of the paradigm (reaction times: t_{86}=0.81, P=0.418, d=0.178 Bayes Factor=0.3; error rates: t_{86}=0.49, P=0.627, d=0.106, Bayes Factor=0.2; Fig. 3.4a). Furthermore, there was a large positive correlation between the validity effects using objects and chunks (R=0.63, CI_{95}=0.40-0.78, P<0.001, Bayes Factor=3499; Fig. 3.4b) suggesting that observers who produced a large validity effect in the chunk version also produced a large validity effect in the object version of the paradigm. These results confirm that classical cueing worked in a very similar manner with objects and chunks.

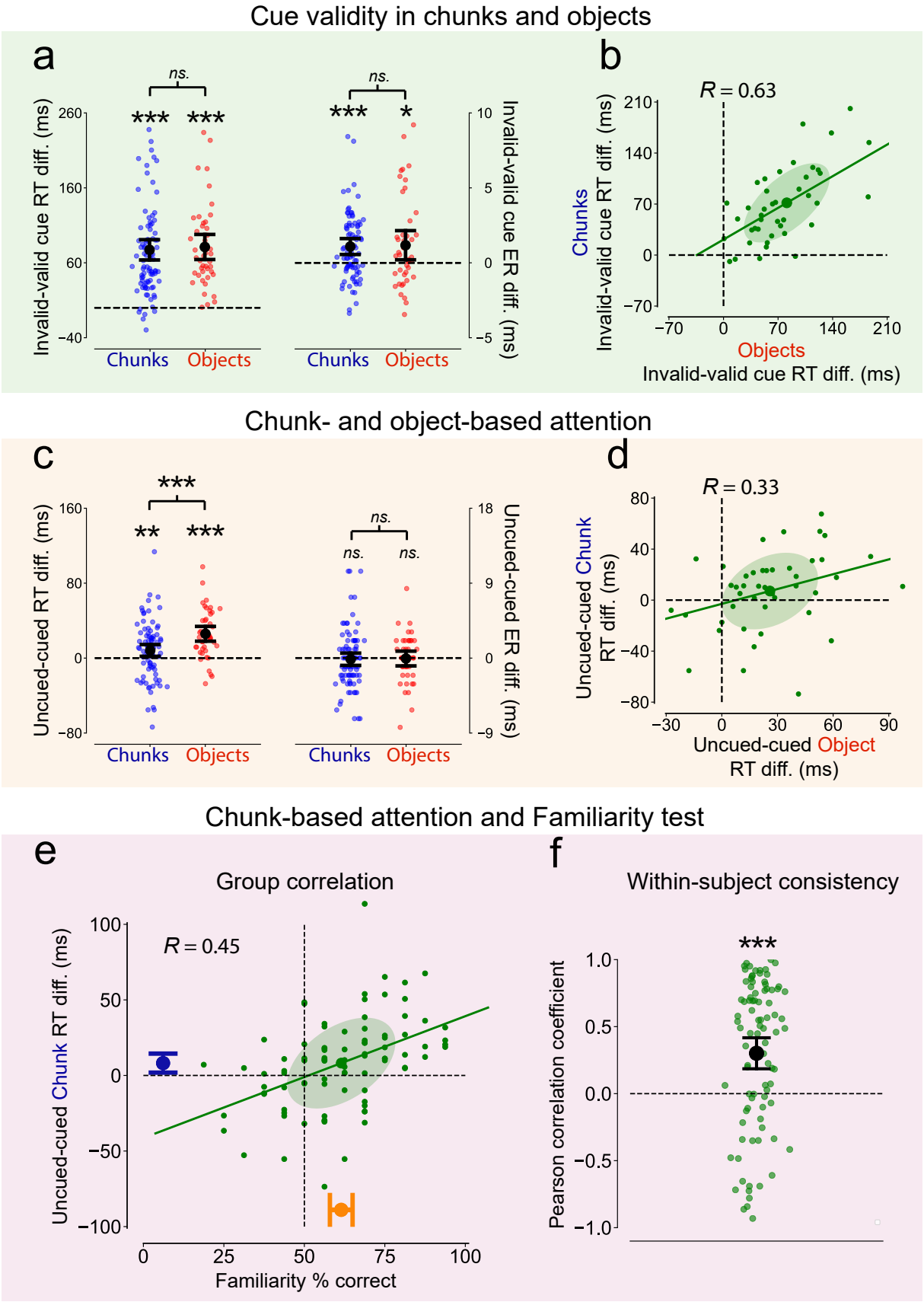


Figure 3.4: Chunk- and object-based attentional effects in Experiment 2. Caption continues on the next page. Adapted with permission from G. Lengyel, M. Nagy, and J. Fiser.

Figure 3.4: (Caption for Fig. 3.4 on the previous page.) **a** The cue-validity effect for chunks (blue) and objects (red). Dots represent the individual observers' validity effect defined as the difference between the median reaction times (right) and mean error rates (left) in the invalid- (uncued) and valid-cue (cued) trials. **b** Correlation between object-based (x axis) and chunk-based (y axis) cue validity with dots representing the corresponding validity effect for each observer. **c** The chunk-based (CBA, blue) and object-based attention (OBA, red) effects. Dots represent the individual observers' OBA/CBA effect defined as the difference between the median reaction times (right) and mean error rates (left) in trials with the target being in an uncued vs. cued chunk/object. **d** Correlation between object-based (x axis) and chunk-based (y axis) attention effects on reaction times with dots representing the corresponding attention effect for each observer. **e** Correlation between the learned statistical structure and the CBA effect with dots in the scatter plot representing each observer's percent correct values in the Familiarity test (x axis, mean in orange) and the extent of their CBA effect (y axis, mean in blue). **f** Within-subject consistency between learning chunks and the evoked CBA effect. Green dots represent the observer's Pearson correlation coefficient between their fraction correct scores and the extent of the CBA effect for each individual chunk. In all plots, error bars denote the 95% confidence intervals of the mean, error ellipses cover one standard deviation, and solid lines represent best-fitting linear regression lines. In the axis labels RT stands for reaction time and ER stands for error rate. n=90 in the blocks with statistical chunks (**a, c, e, f** in blue and green), and n=44 in the blocks with geometric objects (**a, b, c, d** in red and green). Significant differences from zero in **a, c,** and **f** are indicated with ns., P>0.05, *, P<0.05, **, P<0.01, ***, P<0.001, two-tailed paired (difference between uncued and cued or invalid and valid chunk/object trials) t-tests. R values in **b, d** and **e** indicate Pearson correlation coefficients.

Chunk-based attention effect (CBA)

Beyond cue validity, we also successfully replicated the OBA effects reported in earlier studies using objects with visual boundaries (Egly et al., 1994; Moore et al., 1998; Shomstein & Yantis, 2004; Vecera et al., 2000). In the invalid-cue trials, observers responded faster when the target appeared in the cued object albeit not in the cued position (cued-object trials) compared to when it appeared in the uncued object (uncued-object trials) demonstrating the classic OBA effect (t_{43}=6.62, P<0.001, d=1.010, Bayes Factor= $3 \cdot 10^5$, Fig. 3.4c, left panel, in red). More importantly, we found the same pattern of results when statistically defined chunks were used instead of objects with clear boundaries. Observers identified the target faster when it appeared on the cued chunk (cued-chunk trials) compared to when it appeared on the uncued chunk (uncued-chunk trials) demonstrating a clear CBA effect (t_{89}=2.58, P=0.011, d=0.273, Bayes Factor=3, Fig. 3.4c, left panel, in blue). We expected the CBA effect to be smaller than the

OBA effect because the former effect emerges due to chunks implicitly learned in the last half an hour while the latter effect is due to objects based on lifelong learning of visual boundary cues. Indeed, the CBA effect was significantly smaller than the OBA effect (t_{43}=3.84, P<0.001, d=0.586, Bayes Factor=68, Fig. 3.4c). However, there was a significant positive correlation between the CBA and OBA effects (R=0.33, CI_{95}=0.03 - 0.58, P=0.026, Bayes Factor=3, Fig. 3.4d) providing substantial evidence towards a positive relationship between chunk- and object-based attention. A further link could be established between cue validity and OBA by comparing the results in Fig. 3.4b and d. The cue validity effect in Fig. 3.4b indicates the correlation between object- and chunk-based effects for trials where the cue predicted exactly where the target would appear, whereas Fig. 3.4d shows the same correlation for trials where the cue indicates only the correct object/chunk, but not the correct location. The correlation of R=0.63, obtained in the former case, where the object and chunk-based cueing conditions are highly similar, puts an upper bound on how strong the correlation could be in the latter case had the two processes shared exactly the same underlying mechanism. Therefore, the R=0.33 obtained in Fig. 3.4d suggests that chunks and contour-based objects evoke significantly overlapping cognitive processes. There were no similar effects in the error rates either for trials with objects or with chunks (Objects: t_{43}=-0.16, P=0.872, d=0.025, Bayes Factor=0.2; Chunks: t_{89}=-0.42, P=0.671, d=0.045, Bayes Factor=0.1; Fig. 3.4c, right panel).

The consistency of the CBA effect

Next, we tested whether our CBA effect was not just a spurious finding. We found a very significant positive correlation between observers' performance in the Familiarity test, which indicated the extent of their learning, and the size of their CBA biases (R=0.45, CI_{95}=0.26-0.61, P<0.001, Bayes Factor=2833, Fig. 3.4e). To confirm that this strong positive relationship between learning and CBA was not merely due to changes in generic (e.g. alertness-based)

processes, we conducted a within-subject consistency analysis. For each observer, we measured how much the particular chunks they preferred more strongly during the Familiarity test were also the ones that elicited a larger CBA effect. Comparing Familiarity scores and CBA effects for each observer and each chunk separately, we found a very strong and significant within-subject consistency (R=0.27±0.06, t_{89}=4.53, P<0.001, Bayes Factor=941; Fig. 3.4f).

Finally, as in Experiment 1, we measured the CBA effect in the trials in which only two true-pairs were presented to rule out the possibility that the CBA effect emerged only in trials with individual shapes because participants allocated more attention to the true-pairs than to the two individual shapes (see Fig. B.4). We found that the CBA effect was detectable in trials with two true-pairs, and it was significant with the same effect size (t_{89}=2.57, P=0.012, d=0.273, Bayes Factor=3). This again indicates that the chunk-based error rate effect cannot be explained by allocating more attention to true-pairs than to individual shapes per se.

Taken together these results, the chunks learned during VSL elicited a very similar attentional effect to what objects with explicit visual boundaries are known to generate. Furthermore, this chunk-based effect was strongly related to the implicitly learned statistical structure during the VSL, since the stronger a chunk was preferred in the Familiarity test, the stronger attentional effect it evoked in the CBA paradigm. Finally, the correlation between CBA and OBA suggests that related mechanisms could be involved when processing objects or chunks supporting the claim that statistical learning creates object-like representations.

3.2.4 Discussion

The present study provides the first evidence that statistically defined chunks influence visual processes in subsequent search tasks the same way as objects defined by articulated boundary cues do. In the first experiment, observers performed better in a novel 3-AFC visual search

task when the targets appeared on the same chunk as opposed to when the targets appeared on two different chunks. In the second experiment, chunks elicited the same object-based attention effect as was reported in the classical findings of Egly et al. (1994). In both experiments, the chunk-based effect was larger in observers who performed better in the familiarity test that measures the observers' implicit knowledge of the statistical structure embedded in the stimuli. These results have implications in two domains of the research on internal representation in the brain: the nature of object representation and the role of learning in having object representations.

Object representation initially has been approached as a boundary contour problem (Biederman, 1987; Marr, 1982) that later evolved into characterizing a large number of important cues for object formation, such as good continuation (Pizlo et al., 1997; Smits & Vos, 1987), closure (Pomerantz et al., 1977), connectedness (Palmer & Rock, 1994), convexity (Bertamini, 2001; Liu et al., 1999), and regularity of shape (Feldman, 1997, 2000). Here, we argue for a parsimonious integration of these results by stating that the notion of boundary information for the brain is more general than edge contours, and it is based on separating two sets of consistent elements according to some complex statistical measure, which naturally leads to object representations. In the simplest case, these are dark and light local regions giving rise to a luminance boundary or edge. However, apart from such first-order boundaries, there exist for example second-order boundaries that are invisible to mechanisms detecting first-order boundaries, do not necessarily co-occur with the first-order boundaries, and have ecological relevance (Schofield, 2000; Schofield et al., 2010). In addition, there are texture-based, disparity-based or motion-based boundaries (Julesz, 1971) that can be largely independent from luminance-based boundaries and that are more difficult to perceive without prior experience. In this ordering of increasingly abstract examples of boundaries, mid-level visual routines detecting discontinuit-

ies in any arbitrary measure of the stimulus, or Gestalt rules are at an even higher level, the stimuli used in our study reside at the opposite extreme from edges: our elements are grouped and separated based on purely statistical consistencies of co-occurrence without the use of any other low-level visual measure. Yet they evoke the same treatment by our cognition as true contour-based object stimuli do even if only to a smaller extent. Thus, we propose that object representations are defined and object-based effects emerge whenever a sufficient subset of statistical contingencies at various levels of abstraction together indicate a separable entity. We also propose that although objectness seems to be an all-or-none property in most natural settings, in fact, it is a continuum with different degrees of objectness. For example, two solid objects separated by a clear visual gap are perceived as two separate objects until they start to move coherently (Kellman & Spelke, 1983), when they are interpreted as one object with two parts or with a surface marker, and the degree of perceived single-objectness will depend on the level of motion coherence between the two objects.

Regarding the role of learning in forming boundaries and objects by statistical contingencies, a number of earlier results corroborate our proposal that statistical learning leads to object-like representations. Several findings suggest that VSL interferes with perception: it affects the extraction of summary statistics of scenes (Zhao et al., 2011), automatically biases attention (Zhao et al., 2013), modulates perceived numerosity (Zhao & Yu, 2016), creates novel object associations based on transitive relations (Luo & Zhao, 2018) and influences the size perception of the elements within the structure (Yu & Zhao, 2018). Two earlier studies linked perceptual organization and statistical learning between abstract shapes directly (Vickery & Jiang, 2009; Zhao et al., 2014). In Vickery and Jiang (2009) chunks were explicitly delineated from the surrounding with a clear black line, and they found that learning new shape associations with such explicit visual cues led to perceptual grouping. Zhao and colleagues (2014) showed that detect-

ing color change was faster within than across chunks that were defined solely by co-occurrence statistics. Unlike in our paper, observers in that study completed the Familiarity test, in which the true chunks were explicitly shown, before the color change detection task with the chunks, and therefore, they had an explicit memory of the underlying chunks. Nevertheless, these studies provide a partial support to our claim that statistical learning has a key role in the emergence of object representations in humans.

Another support for the crucial role of learning in forming object representations comes from infant studies. Automatic VSL has been demonstrated amply across various modalities not only in adults but in infants as well (Fiser & Aslin, 2002; Quinn & Bhatt, 2005; Saffran & Kirkham, 2018), suggesting that infants and adults are equally capable of learning the co-occurrence statistics of scenes (Aslin, 2017). Infants are also known to segment and represent objects initially only by a subset of the available sensory cues, the most important cues being surface motion and arrangement, while their ability to utilize the other cues, such as Gestalt rules or smooth contours develops later (Spelke, 1990). This gradual incorporation of more complex cues by infants (Bertenthal, 1996; Spelke, 1990) is compatible with the idea that statistical learning mechanisms have a key role in the emergence and elaboration of object representation during infancy. Further support comes from another line of infant studies demonstrating that prior experience with given objects together or separately brings forward the time when the infant is able to perform object segregation properly with the particular objects (Needham, 1997; Needham & Baillargeon, 1998; Needham & Modi, 1999). While these results are strongly suggestive, future studies will be required to test precisely the relative importance and limits of statistically learned vs. innately available cues in object representation across ages.

Our results show only correlation between the measured object-based effects and the amount of learning, thus we cannot completely exclude the possibility that the co-variation is due to a

common source, and learning contingencies is not causally linked to the emergence of object-like representation of the input. An alternative interpretation of our results could be that object-based attention is not really object-based, and objects and chunks share this kind of attentional effect, which should be properly called an "object-and-chunk-based" attention. However, this is unlikely for two reasons. First, the correlation remained strong after controlling general improvement in performance, and this reduces the probability of an uncovered common cause since assuming a dynamically strengthening hidden cause that is related neither to general performance nor to learning contingencies is implausible. Second, there exists no visual cue in our chunk stimuli other than statistical contingencies that would selectively map to the features that were implied as causes of OBA in objects, while the features that were implied (long contours, similar textures/colors, Gestalt structures, etc.) all represent strong examples of statistical contingencies. Therefore, based on parsimony, we propose that the emergence of the chunk-based advantage in Experiment 1 and the chunk-based attention in Experiment 2 are direct consequences of implicitly collecting enough statistical evidence by VSL to treat the chunks as a preliminary objects, and automatically initiating object-related processes on them. Clearly, this does not mean that the object-like representation emerging after a brief VSL can be considered as fully-blown, real mental objects, as these preliminary object-like representations need to be fortified by further experiences to pass several additional criteria to reach the representational richness of true mental objects. Whether and under what conditions VSL mechanisms can produce such fully developed mental object representations needs to be clarified by future studies.

Earlier computational studies can point to possible computations showing how statistically defined chunks and objects are related (Fiser & Lengyel, 2019; Orbán et al., 2008; Perruchet, 2018). When observers are faced with an unfamiliar environment with unknown statistical

structure composed of shapes, they learn and compress the information about the stimuli in terms of meaningful latent chunks from the shapes instead of representing only recursive pairwise associations between those shapes (Orbán et al., 2008; Perruchet, 2018). Therefore, we argue that these latent chunks extracted hierarchically based on the statistical regularity in the sensory input are the building blocks of object-like representations. Investigating visual scenes with low-level features, a recent study provided a computational framework, based on hierarchical Bayesian clustering, that demonstrated how an image can be represented by mixture components organized hierarchically, and how such representations can capture most Gestalt rules through probabilistic inferences (Froyen et al., 2015). According to the main proposal of this thesis (see section 1.4), such hierarchical chunk-representations, using probabilistic learning, that makes inferences across multiple levels simultaneously can also link VSL -and therefore object representations- to low-level perceptual effect and perceptual learning (Fiser & Lengyel, 2019).

Regarding the neural correlates of object-based perceptual effects, an fMRI study reported that in the early visual cortex, visual error predictions spread between the parts of the same object (Jiang et al., 2016). This suggests that already in the early visual cortex, the context for computing the prediction error is defined by the objects rather than by low-level visual cues. If this is correct, early visual areas should also manifest increased gamma synchrony with higher areas similarly to what has been reported in relation to object-based attentional effects between the inferior frontal junction and the fusiform face and parahippocampal place areas (Baldauf & Desimone, 2014). Moreover, we posit that this effect should increase with learning the underlying chunk-structure of an unknown visual stimulus.

In conclusion, the present results provide a significant step toward linking the concept of object representations to implicit statistical learning of environmental structures through rede-

fining the fundamental requisites necessary for the perception of a new object.

3.3 Study 3 - Across-modality generalization

3.3.1 Introduction

In the previous study we demonstrated that chunks of abstract shapes defined by co-occurrence statistics elicited very similar attentional and perceptual processing to what true objects defined by visual boundaries elicited. This suggests that participants built abstract, object-like representations during SL and these representations (the abstract chunks) served as organizational units for attention allocation and other perceptual processes in the subsequent visual search task. Therefore, during SL participants learned to segment our environment into meaningful, object-like units/chunks. However, it is still unknown how abstract these representations about the chunks are?

The level of abstractness of true object representations exceeds view-invariance across all viewing conditions in the visual domain, and it also includes amodal representation of objects across all sensory modalities under all "possible" cross-conditions. This coherent organization of information across different modalities is crucial for efficiently interacting with the world and lies at the heart of the concept of what defines an object (Amedi et al., 2001; Pascual-Leone & Hamilton, 2001; Streri & Spelke, 1988).

In the second study of this chapter, we investigated whether the representation of the chunks built during SL are also abstract and amodal similar to real object representations. Considering the visual and the haptic modalities, we hypothesized that participants should be able to predict haptic properties of objects based on just visual statistics, without any specific prior haptic experience with them and vica versa, they should be able to predict visual properties based on

haptic statistical exposure alone without receiving any form of feedback fostering cross-modal generalization. " Inspired by the concept of “one-shot learning” in Machine Learning, which refers to the ability of generalizing to a new context after only one testing example, we refer to this generalization across modalities without any testing example as “ zero-shot" generalization (c. f. Fu et al., 2014; Lampert et al., 2009).

We used the same set of artificial stimuli as in the previous study, in which the statistical contingencies defining objects had, by design, no correlation with boundary cues. This avoided the problem that, under natural conditions, boundary cues and edges can be correlated with the statistical contingencies of objects (Geisler et al., 2001). We created an inventory of artificial “objects”, such that each object was defined as a unique pair of unfamiliar shapes (Fig. 3.5A, inventory, colouring and gaps within pseudo pairs for illustration only). Note that only the individual shapes, but not the pairs defining the objects of the inventory, had visible boundaries. Therefore, boundary cues were uninformative with regard to the object identities, and instead participants could only rely on the statistical contingencies among the shapes that were created in either the visual or the haptic modality during an exposure phase. We then examined how the information extracted from the visual or haptic statistics affected performance on both a visual familiarity and a haptic pulling test, thus measuring within-modality learning as well as across-modality generalisation of statistical information.

In two experiments we found clear evidence towards within-modality learning and "zero-shot", across-modality generalization when participants were exposed to visual statistics alone and visual-to-haptic generalisation were measured (Experiment 1), and also when participants were exposed to haptic statistics alone and haptic-to-visual generalisation were measured (Experiment 2).

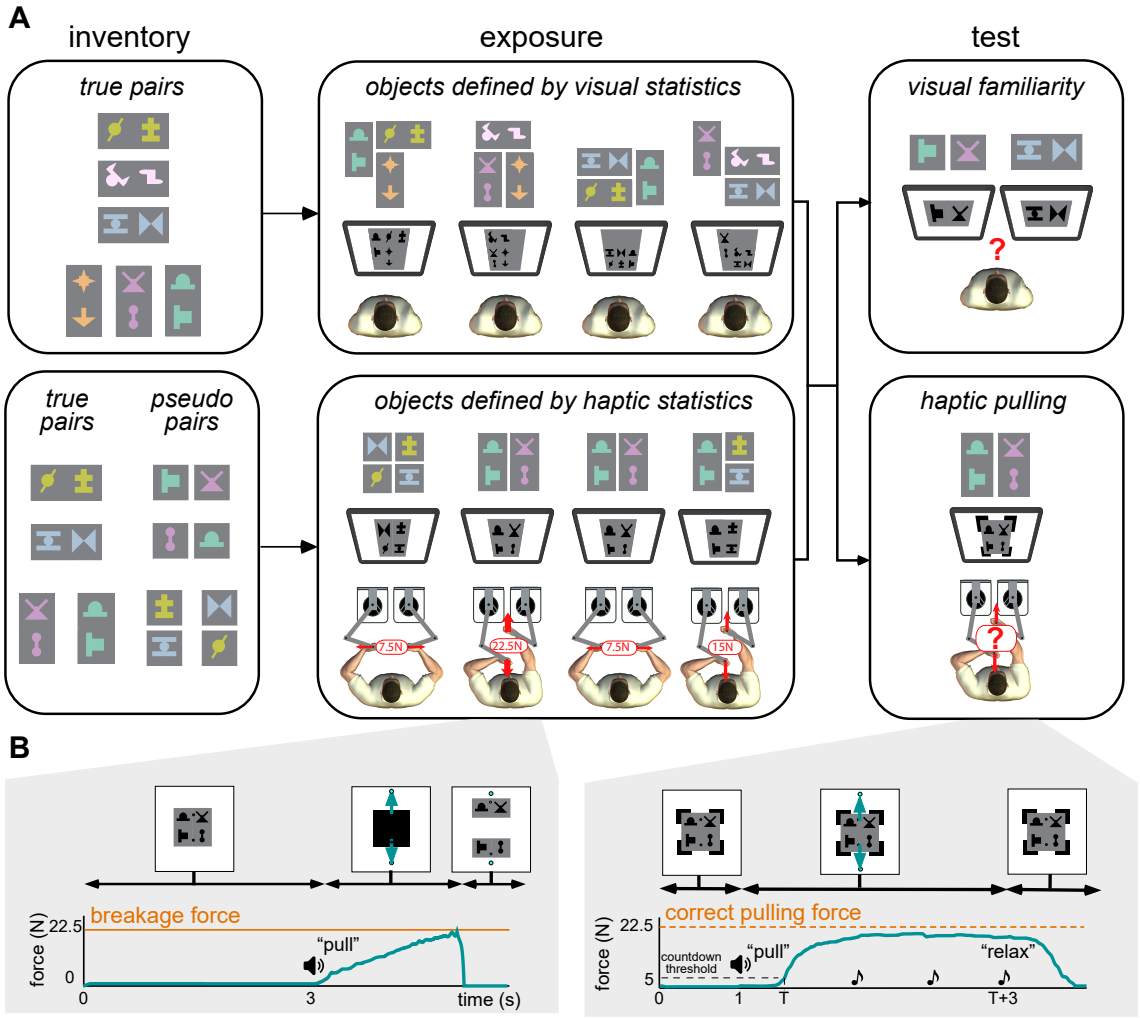


Figure 3.5: Experimental paradigm. **A.** Main phases of the experiments. **Left.** An inventory was constructed by arranging abstract shapes into horizontal and vertical pairs. True pairs behaved as objects: their shapes always appeared together, and in the same relative spatial configuration, and were hard to pull apart physically. Pseudo pairs served as controls: they had consistent visual statistics but were as easy to pull apart as two separate objects (indicated by the small separation between their shapes). Colouring and separation for illustration only, participants saw all shapes in grey-scale during exposure and testing, with no gaps between them, so that no visual cues separated the pairs of a compound scene (as shown on screens in the center and right panels). **Center.** During the exposure phase, participants experienced a sequence of visual scenes showing compound objects consisting of several pairs. The way the image displayed on the screen was constructed from the inventory is shown above each screen in colour for illustration. In the first experiment (top), participants observed compound scenes each constructed from three true pairs of the inventory. Caption continues on the next page. Adapted with permission from G. Lengyel, G. Žalalytė, A. Pantelides, J. N. Ingram, J. Fiser, M. Lengyel and D. M. Wolpert.

Figure 3.5: (Caption for Fig. 3.5 on the previous page.) In the second experiment (bottom), on each trial, a compound scene consisting of two pairs (true or pseudo) was displayed and participants were required to pull the scenes apart in one of two directions as shown. A bimanual robotic interface (Howard et al., 2009) was used so that participants experienced the force at which the object broke apart (breakage force shown in red) but, crucially, visual feedback did not reveal the identity of true and pseudo pairs (see Methods). Thus, only haptic information distinguished the true and pseudo pairs as the force required depended on the underlying structure of the scene. **Right.** In both experiments, participants finally performed two tests. First, in the haptic pulling test (bottom), participants were asked to pull with the minimal force which they thought would break apart a scene, composed of true or pseudo pairs (in both directions). We measured this force by "clamping" the scene so that no haptic feedback was provided about the actual breakage force (black clamps at the corners of the scene). Crucially, the visual display also did not reveal the identity of true and pseudo pairs. Second, in the visual familiarity test (top), participants were asked to select which of two scenes presented sequentially appeared more familiar. One scene contained a true pair and the other a chimeric pseudo pair. Selecting the true pair counted as a correct response, but no feedback was given to participants as to the correctness of their choices. **B.** Timeline of events in haptic exposure and test trials (displayed force traces are from representative single trials). **Left.** Haptic statistical exposure trials had scenes consisting of combinations of true and pseudo pairs of the inventory (top). After a fixed amount of time, the scene was masked (black square covering the scene), then pulling was initiated ("pull" instruction was played), and the scene was unmasked and shown as separated once the pulling force (green arrows and curve) exceeded the breakage force (orange line). **Right.** In the haptic pulling test, participants were asked to generate a pulling force which they thought would be just sufficient to break the scene apart (ideally the breakage force corresponding to the scene, orange dashed line). The scenes were constructed using the pairs of the inventory without any visible boundary between them and held together by virtual clamps at the corners of the scene (top). Pulling was initiated ("pull" instruction), and once the participant's pulling force (green arrows and curve) exceeded a 5 N threshold (dashed black line), three beeps were played at 1 s intervals (notes). The clamps remained on until the end of the trial (top), so the scene never actually separated, and after the third beep (at which the pulling force was measured) participants were asked to "relax". See sections 3.3.2.1 & 3.3.3.1 in the Methods for details of the variant used in the haptic exposure task.

3.3.2 Experiment 1 - Visual-to-haptic generalization

First, we examined visual learning and visual-to-haptic generalisation. During exposure, participants (N=20, after exclusion, see section 3.3.2.1, Methods) experienced a sequence of visual scenes, each consisting of a spatially contiguous cluster of 6 shapes displayed on a grey square (Fig. 3.5A, exposure, top) very similar to the scenes in the previous study (Fig. 3.1A). Unknown to the participants, each 6-element scene was constructed by combining three of the objects from the inventory of true pairs (coloured explanatory diagrams shown above displays). Therefore, the objects could only be identified based on the consistent visual co-occurrence

of their constituent shapes across scenes as participants did not have any experience with the scenes' haptic properties.

After the exposure phase we tested participants' within-modality statistical learning performance on a visual familiarity test in which they had to choose which of two pairs in a trial was more familiar: a true pair or a "chimeric" pseudo pair constructed of two shapes belonging to two different true pairs of the inventory (Fig. 3.5A, test, top). This test is analogous to comparing familiar scenes that contain real-world objects (e.g. rabbits or deers), and thus comply with the known statistical regularities of the world, with unfamiliar scenes containing chimeras (e.g. a wolpertinger — a mythical hybrid animal with the head of a rabbit, the body of a squirrel, the antlers of a deer, and the wings of a pheasant; contributors. (2018)).

Critically, we also tested whether the exposure to visual statistical contingencies also generalised to participants' judgements as to the force required to pull apart novel compound objects. In order to provide participants with general experience about the forces associated with pulling objects apart in different configurations in our set-up, but without any reference to the objects of the shape-inventory, we pre-trained them on a task that required them to pull apart scenes consisting of coloured rectangles as objects which thus had clear boundaries (Fig. 3.6). Participants then performed the main pulling task which used the shapes of the inventory, such that each scene consisted of two true pairs of the inventory, arranged as a 2×2 square (Fig. 3.5A, haptic pulling test). On each trial, participants had to pull on a scene in a predetermined direction with the minimal force they thought was necessary to separate the scene (into two vertical pieces for horizontal pulling and vice versa; Fig. 3.5B, right). Crucially, we simulated clamps at the corners of the scene that prevented it from actually separating, so that participants received no haptic or visual feedback as to whether they exerted the correct amount of force, and thus their performance must have been solely based on what they had learned about the visual stat-

istics of the objects during the exposure phase. Specifically, given the pre-training, and their knowledge of the objects of the inventory, participants were expected to pull harder when the pulling direction was parallel to the orientation of the pairs as this required both objects to be broken in half. Conversely, we expected them to pull less hard on trials in which the pulling direction was orthogonal to the orientation of the pairs as this only required them to separate the two objects from each other. We measured participants' performance as the correlation (ρ) between their pulling force and the required breakage force (see Fig. 3.8).

We found that participants were able to build representations about true pairs after the visual exposure and they preferred the true pairs over the pseudo pairs in the familiarity test. Furthermore, participants pulled harder when they had to break true pairs apart into their constituent shapes compared to when they had to separate two true pairs from each other demonstrating that participants generalized the visually learnt pairs to the haptic domain as units/chunks that stick together in a similar way as parts of an object would stick together.

3.3.2.1 Methods

Participants

In the visual statistical exposure experiment, 28 participants (age range 19-39, mean 25, 20 women) gave informed consent and participated. Eight participants were excluded from full analysis as they did not achieve significant performance in haptic task training (see section 3.3.2.1, Exclusion criteria). The final sample therefore comprised of 20 participants (age range 20-39, mean 25, 15 women). Data was collected in two installments. First, we made a preliminary estimate of the approximate number of participants we would need for significant results and collected data accordingly. This resulted in 16 participants after the exclusion criteria were applied. All our main results (relationship between visual and haptic performance in each ex-

periment) were highly significant and resulted in Bayes factors>10. Subsequently, an external expert not involved either in the design of the study or in the analysis of the data, or invested in the success of our study, suggested that data from 20 participants in each experiment should be collected. Therefore, we collected data from additional participants to reach 20 participants after exclusion in each experiment. Again, all our main results remained highly significant. Thus, the process of adding participants, and the consistent usage of Bayes factors throughout (see below), ensured our study was not biased towards favorable results (Dienes, 2011). All experimental protocols were approved by the University of Cambridge Psychology Ethics Committee.

Equipment

We used two vBOTs to provide haptic stimuli and record haptic responses (during haptic exposure and testing, respectively). The vBOT is a custom-built back-drivable planar robotic manipulandum exhibiting low mass at its handle, with the handle position measured using optical encoders sampled at 1000 Hz, and torque motors allowing endpoint forces to be specified (Howard et al., 2009). Participants were seated in front of the apparatus and held one vBOT handle in each hand. By using two horizontally adjacent vBOTs, we applied haptic stimuli and recorded responses bimanually (Fig. 3.5A, Haptic exposure, and Haptic testing). Visual stimuli were displayed using a computer monitor projected to the participant via a horizontal mirror mounted above the vBOTs (Fig. 3.5A, Haptic exposure, Visual exposure, Haptic testing, and Visual testing). During haptic exposure and testing, the participants' veridical hand positions were represented using two cursors (0.3 cm radius) overlaid in the plane of the movement. Responses during visual testing were recorded by closure of the switches on the vBOT handles.

Visual stimuli

In both experiments, following previous work (Fiser & Aslin, 2001), visual stimuli for statistical learning consisted of 12 (visual statistical exposure experiment) or 8 (haptic statistical exposure experiment) black abstract geometric shapes of intermediate complexity arranged along a regular grid (without grid lines shown) on a grey background (Fig. 3.5A). Unbeknownst to participants, the shapes were grouped into "true pairs" (the "objects"), such that constituent shapes of a pair were always shown together and in the same spatial (horizontal or vertical) arrangement, and each shape was part of only one true pair (Fig. 3.5A, Inventory, True pairs, coloured only for illustrating pair identity).

The inventory of shapes that could be used for constructing visual scenes included three horizontal and three vertical (visual statistical exposure experiment) or two horizontal and two vertical such true pairs (haptic statistical exposure experiment) as well as an equivalent number of "pseudo pairs". The pseudo pairs re-used the shapes of the true pairs such that each horizontal (vertical) pseudo pair consisted of two shapes belonging to two different vertical (horizontal) true pairs, one of them being the top (left) the other the bottom (right) shape of its original true pair (to avoid accidental constellations that could have appeared using true pairs), and each shape was part of only one pseudo pair (Fig. 3.5A, Inventory). The assignment of shapes to true and pseudo pairs was randomized across participants to control for effects due to specific shape configurations.

During visual exposure, only true pairs and no pseudo pairs were shown. During haptic exposure, pseudo pairs were displayed with the same visual statistics as true pairs (but they differed in their haptic properties, see below). Each visual scene during exposure (and haptic testing) contained several pairs (three for visual exposure, and two for haptic exposure and testing) in juxtaposition, in a non-occluding manner, without any border lines separating them.

Each visual scene during the visual familiarity test consisted of a single true or pseudo pair.

In general, note that the co-occurrence statistics relevant for statistical learning of pairs included both the number of times two shapes appeared together and the number of times each appeared alone (see a formal definition in (Orbán et al., 2008)). This meant that true pairs had stronger overall statistical contingencies in the visual than in the haptic statistical exposure experiment as shapes of a true pair never appeared without each other in the former while they did in the latter due to pseudo pairs.

Controlling for special cues

Crucially, the instructions to the participants did not refer to the existence of objects in any way, only that there were visual or haptic patterns they needed to observe (see also below). We controlled the stimuli for low-level visual segmentation cues, such that there were no boundaries or colour differences revealing the objects present in a scene (the colour coded shapes in Fig. 3.5A illustrate the construction of scenes but these were never displayed to participants). Moreover, while individual shapes were clearly separated, the separation between adjacent shapes belonging to the same or different objects was the same and thus uninformative as to object boundaries. Therefore, objects were not defined, as might naively be expected from observing individual scenes, at the level of a single shape or all shapes in a scene. Instead, the only information available to identify the objects was the statistics of either their visual co-occurrences or of the physical interactions they afforded across the exposure scenes.

Visual statistical exposure experiment

The experiment consisted of four phases: (1) visual statistical exposure, (2) haptic task training, (3) haptic pulling test, and (4) visual familiarity test.

Visual statistical exposure: Participants were exposed to a series of scenes constructed using an inventory of 12 shapes arranged into 6 true pairs, each scene being composed of 3 pairs arranged along a 3-by-3 grid (Fig. 3.5A, Visual exposure, top coloured pairs shows the construction and below the screen display, see also above). To ensure there were no obvious spatial cues identifying individual pairs, the positions of the pairs within the grid were randomized with the following constraints: (1) at least one shape in a pair had to occupy a grid location adjacent to one shape in another pair, (2) the central square needed to be occupied by a shape, and (3) the exact same configuration of 3 pairs but at a different location on the grid were considered the same. These spatial constraints generated a set of 444 unique scenes, in which each of the 6 pairs appeared 222 times (see Fig. 3.6). The scenes were generated as follows (where H and V are horizontal and vertical pairs, respectively):

- 3H gives 3 (identity of H1) × 2 (identity of H2) × 1 (identity of H3) × [1 (all aligned) + 2 (left/right displacement of one H) × 3 (which H is displaced)] = 42

- 2H & 1V with the two H aligned gives 3 (identity of H1) × 2 (identity of H2) × 3 (identity of V) × 2 (location of V on left/right) × [1 (V aligned) + 2 (V offset top/bottom)]= 108

- 2H & 1V with the two H offset gives 3 (identity of H1) × 2 (identity of H2) × 2 (location of H1 offset left/right) × 3 (identity of V) × 2 (location of V above/below) = 72

The total number of unique scenes are 222, and the same for 3 V and 2V & 1H, giving a total of 444 unique scenes. These scenes were presented in a pseudorandom order one at a time for 700 ms, with 1-s pauses between them, and participants were instructed to simply view the scenes without any explicit task other than paying attention so that later they could answer simple questions related to the scenes (Visual familiarity test). The instructions simply asked participants to “observe each display carefully so that you can answer simple questions related

to the pattern of symbols that you observed in all of the displays”.

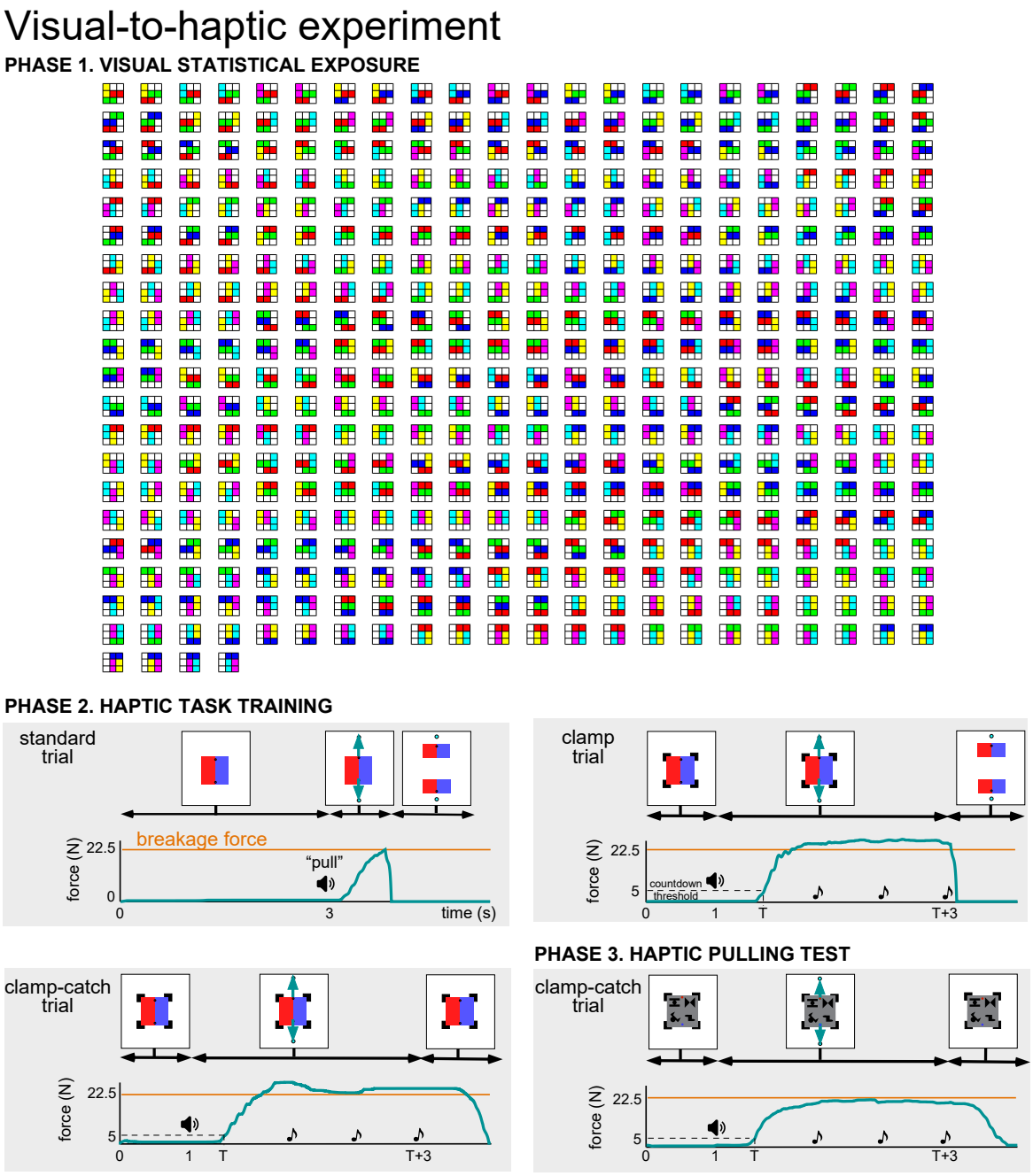


Figure 3.6: Phases of the visual statistical exposure experiment. **Phase 1.** Visual statistical exposure consisted of 444 scenes (shown in a pseudorandom order), each using a combination of 3 true pairs of the inventory (see Fig. 3.5A). Each true pair is shown here as a uniquely colored 2×1 block for illustrative purposes only and was replaced by true pair shapes in the experiment (assignment of shapes to colors was randomized across participants). The grid lines were not displayed in the experiment. Caption continues on the next page. Adapted with permission from G. Lengyel, G. Žalalytė, A. Pantelides, J. N. Ingram, J. Fiser, M. Lengyel and D. M. Wolpert.

Figure 3.6: (Caption for Fig. 3.6 on the previous page.) **Phase 2.** Haptic task training consisted of two 2×1 rectangular coloured blocks (insets), which needed to be pulled apart in both the horizontal and vertical (shown) direction. On standard trials (upper left), participants were instructed to "pull" (green arrows show applied forces, not movement) after the scene had been displayed for 3 s (black speaker icon), and the scene separated when the pulling force (green trace) exceeded the breakage threshold (orange line). The force was low for separating the block along their boundary and high when separating both coloured blocks into two (shown). On clamp trials (upper right) participants were asked to generate a pulling force which they thought would be just sufficient to break the scene apart. Clamps held the objects together initially. Once the participant's pulling force exceeded a 5 N threshold (dashed line), three beeps were played at 1 s intervals, and the clamps were removed on the third beep. At that point the scene broke apart if the force exceeded the threshold, otherwise the participant had to increase their pulling force to break the scene. On clamp-catch trials (lower left) the clamps remained until the end of the trial and participants were asked to relax after the third beep. **Phase 3.** In the haptic pulling test (lower right), clamp-catch trials were used with scenes that were constructed using the (true) pairs of the inventory without any visible boundary between them. **Phase 4.** The visual familiarity test consisted of 72 trials (not shown, see Fig. 3.5A). In each trial, two scenes were displayed, one including a single true pair, the other (order counterbalanced) including a pseudo pair, and participants were asked to select the one that appeared more familiar to them. Green traces in 2 and 3 show the time course of the pulling force on representative trials for a single participant.

Haptic task training: Before haptic testing, participants completed a haptic training task on the vBOT in order to familiarise them with the forces associated with pulling apart objects in different configurations (Fig. 3.6). Each scene consisted of two 2-by-1 rectangles (the "objects", both being either horizontally or vertically oriented) touching on their long sides, so that the configuration was a 2-by-2 block of coloured pieces. In order to avoid any ambiguity about object boundaries in this case, the identity of the two rectangles was clearly revealed by the different colours of the two rectangles (four colours were used in total). After each scene appeared on the screen, the two vBOTs moved the participant's hands passively to circular placeholders on opposite sides of the scene (either vertically or horizontally, chosen randomly). After a period of time (3-s for Standard trials and 1-s for other trial types, see below) the participant was instructed to pull the object apart (computer generated speech "pull") in this predetermined direction. Haptic feedback was provided by simulating a stiff spring (spring constant 30 N/cm) between the handles with a length set to 16 cm corresponding to the initial

hand separation (see below). On the next trial, the pulling procedure was repeated with the orthogonal pulling direction with the same scene, after which the next scene was generated. The training consisted of three trial types: Standard, Clamp and Clamp-catch trials.

On Standard trials, as participants increased their pulling force against the simulated stiff spring, the object broke apart both haptically (force reduced to zero) and visually (split in the direction corresponding to the pull direction) at a predetermined force threshold. Crucially, the threshold at which the scene broke apart depended on its configuration, and in particular whether the pre-set pulling direction required the breaking of objects (pulling direction parallel to the orientation of the objects, and hence to the boundary between them) or not. Specifically, the threshold pulling force was determined for each scene by simulating forces between individual pieces such that pieces belonging to the same object were attached by 11.25 N, and pieces belonging to different objects were attached by 3.75 N. This meant that pulling two objects apart without breaking them required a low force (7.5 N) whereas breaking each object into two required a high force (22.5 N) (Fig. 3.6).

Clamp trials were identical to Standard Trials except that the objects were held together initially by virtual clamps displayed at the four corners. Once the participant started to pull (pulling force exceeded 5 N), three tones were played at 1-s interval and participants were asked to generate the minimal force which they thought would break the scene apart by the final tone. The clamps then disappeared and the scene separated if the force threshold was exceeded. Otherwise, participants were instructed to increase their pulling force until the scene separated.

Clamp-catch trials were similar to Clamp trials except that after the final tone the clamps remained and participants were instructed to "relax" (stop pulling) so that the scene did not actually break apart on these trials and no feedback on the accuracy of the participant's pulling force was given.

Participants were exposed to a total of 56 trials: 24 Standard trials, followed by 16 Clamp trials, and finally 16 Clamp-catch trials.

Haptic pulling test: This test followed a similar format to the haptic task training but using the true pairs of the original shape-inventory as objects, which thus had no visible boundary between them, and only Clamp-catch trials (i.e. no feedback on the accuracy of their pulling force was ever given, and no scenes were ever separated, see above). Visual scenes with a 2×2 block of four shapes were displayed such that two pairs with the same orientation (both horizontal or both vertical) were chosen randomly from the set of all true pairs. Participants were presented with 48 scenes in total (2×24-trial blocks). Within each 24-trial block each combination of two true pairs of the same orientation was presented twice, once for horizontal and once for vertical pulling (the order of scenes was randomly permuted within each block). Note that while this phase did not provide haptic experience with the objects, it did provide additional visual statistical information, in somewhat simpler scenes (2 rather than 3 objects in each) but still without boundaries, so for these purposes it could be regarded simply as extra visual familiarisation.

Visual familiarity test: Lastly, participants performed a sequence of two-alternative forced choice trials. In each trial, they had to indicate which of two consecutively displayed scenes was more familiar. Scenes were presented sequentially for 1-s with a 1-s pause between them. One of the scenes contained a true pair, the other a pseudo pair of the same orientation. Horizontal pseudo pairs were generated from the shapes of vertical true pairs while the vertical pseudo pairs were generated from the shapes of horizontal true pairs. Participants selected which pair was more familiar by closing the switch on the left (1st pair) or right (2nd pair) vBOT handle. Participants performed 72 trials (2×36-trial blocks) in total. Within every 36-trial block each

true pair was compared to each pseudo pair of the same orientation in each order exactly once (the order of trials was randomly permuted within each block). Note that only in this last phase did participants see individual, separated objects (constructed from the shapes of the inventory), of which the boundary was thus obvious.

Debriefing

Occasional perfect (100%) performance on the visual familiarity task and informal debriefing with the first batch of participants suggested that some might have been developing explicit knowledge of the pairs. Therefore, we chose to perform quantitative debriefing for the final 23 participants at the end of the experiment (16 in the visual and 7 in the haptic statistical exposure experiment). Participants were asked "Did you notice anything about the shapes during the exposure phase of the experiment?". If they said "yes" then they were asked "What was it that you noticed about the shapes?" and if they said something about pairs they were shown the inventory of shapes separated on a page and instructed: "Point to all the shapes that form part of pairs that you remember." Participants were free to indicate as many pairs as they wanted, and if they identified less than the number of true pairs in the inventory they were not required to guess the remaining pairs. The 8 participants who did not notice any pairs were given an explicitness score of 0, while the other 15 participants correctly identified at least one pair and were given an explicitness score equal to the number of correctly identified pairs divided by the total number of true pairs in the inventory. Thus, the score was only based on correctly identified pairs and we ignored incorrect pairs so as to err on the side of increased explicitness in our measure.

Note that this measure of explicit learning not only required that participants had explicit knowledge of the pairs but also that they had an explicit "meta-cognitive" sense for this knowledge. It could have been possible that some would have identified some pairs even without having an explicit sense that they did, but note that our visual familiarity task already tested their

knowledge of pairs by a two-alternative forced choice familiarity judgment (typically taken as an index of implicit learning) and this additional debriefing at the end of the experiment instead served to rule out that highly cognitive operations accounted for all across-modality generalization.

Data analysis

Basic performance measures: Familiarity trials provided binary data, in which choosing the true pair counted as a correct response. As a summary measure of familiarity, we calculated the fraction correct across all trials for each participant. In haptic trials, we recorded the position and force generated by the vBOTs at 1KHz. Responses in pulling test trials provided the pulling force generated by participants on the final beep after 3 seconds (Figs. 3.6 and 3.9). The vBOTs are limited to generating a maximum pulling force of 40 N and therefore pulling forces were clipped at 40 N and this happened on 0.21% of both haptic clamp trials in the visual and haptic statistical exposure experiments, respectively. As a summary measure of the pulling test performance, we calculated the correlation (ρ) between the pulling force and the breakage force across all trials. This measured how much their pulling force aligned with the required breakage force while being insensitive to an overall mismatch in the scale or offset of forces. The only critical feature for our hypothesis was that participants should pull harder to separate true pairs into two, compared to pseudo pairs or junctions between pairs, and the correlation measure with breakage force reflects this feature. (Similar results, not shown, were obtained by using the slope of the correlation instead, which takes into account the scale of forces, but remains insensitive to the reliability with which participants generate their forces.) Even though in the first experiment (objects defined by visual statistics), only two levels of breakage force were possible, we still used correlation to keep our results comparable with the second experi-

ment (with three levels of breakage force). Nevertheless, note that in this case, the correlation, ρ, was also monotonically related to the sufficient statistic, t, that a direct a comparison (t-test) of the pulling forces at the two breakage force levels would have used (with the same number of trials at each): $t^2 \propto \frac{\rho^2}{1-\rho^2}$.

Participants' performance on the haptic pulling and the visual familiarity tests were compared across the two experiments with t-tests. In both generalization tests (haptic pulling test in the visual statistical exposure experiment, and visual familiarity test in the haptic statistical exposure experiment) participants completed two blocks of the same test trials (in a different randomization, see above). In order to test whether there was a significant change in performance throughout the test trials, we compared the performance in the first and the second block using a paired t-test.

The rectified exponential-binomial model: For each experiment, we fit a rectified exponential-binomial model to predict participants' visual familiarity performance (fraction correct, fc) from their haptic pulling performance (correlation, ρ). This model was not intended to be a mechanistic model of how participants solved the tasks but as a phenomenological model capturing the main aspects of the data. Specifically, it captured three intuitions given our hypothesis that behaviour on the two tasks was driven by the same underlying representation of objects. First, performance on both tasks should depend on how well a participant acquired the inventory of objects through experience in the exposure phase of the experiment, and this common cause should cause co-variability with a monotonically increasing (positive) relationship between the two performance measures. As fc is upper bounded at 1, we chose a saturating exponential function to parametrise this relationship. Second, participants with chance or below-chance haptic performance ($\rho \leq 0$) should have learned nothing about the objects and therefore would have a baseline visual familiarity performance which is independent of ρ. This baseline could in

principle be above chance, especially in the visual statistical exposure experiment where participants learn the visual statistics but do not generalise to the haptic domain. Third, performance on individual trials was statistically independent, given the strength of the object representation of the participant. The rectified exponential-binomial is a three-parameter model that captures these intuitions:

$$f_{\mathrm{c}} = \frac{T_{\mathrm{c}}}{T}, \text{ where } T_{\mathrm{c}} \sim \mathrm{Binomial}(P(\rho), T), \text{ and } P(\rho) = \begin{pmatrix} \beta_0 & \text{if } \rho \leq 0 \\ \beta_1 + (\beta_0 - \beta_1)\, e^{-\rho/\lambda} & \text{otherwise} \end{pmatrix} \tag{3.1}$$

where T_{c} is the number of correct and T is the total number of trials in the visual familiarity test (T=72 and T=32 for the two experiments, see above), β_0 and β_1 determine the range of f_{c}, and λ controls the rate of rise of the exponential. We used a likelihood ratio test to examine the null hypothesis that there was no relation between fraction correct and correlation H0: $\beta_0 = \beta_1$ and thus λ has no effect. In order to compute confidence intervals around the maximum likelihood fits (solid red lines in Fig. 3.7), we used the "profile likelihood" method (Venzon & Moolgavkar, 1988). That is, the $1 - \alpha$ confidence region encloses all parameters values for which the log likelihood is within $\chi^2_{1-\alpha}(n)/2$ of the maximum log likelihood, where N is the number of parameters being estimated via the method of maximum likelihood (Appendix A in McCullagh and Nelder (1989)). Briefly, we sampled 100,000 parameter sets from the Laplace approximation of the log-likelihood (i.e. a Gaussian approximation centred on the maximum likelihood parameter set, with the inverse covariance determined by the local Hessian of the log-likelihood; Bishop (2016)) We rejected those samples for which the negative log-likelihood fell from the maximum by more than $q/2$ where q was the 95th quantile of the χ^2 distribution with 3 degrees of freedom. We then estimated the 95% confidence of the maximum likelihood fit as the extrema of the predictions obtained from the remaining parameter set samples (shaded

red regions in Fig. 3.7).

We also computed the Bayes factor to directly compare the two hypotheses: (1) that there was a systematic relationship between visual and haptic performance as predicted by the rectified exponential-binomial (Eq. 3.1), and (2) the null hypothesis, that is that there was no relation between visual and haptic performance (see also above). This was computed as the ratio of the (marginal) likelihoods of the two hypotheses each of which was approximated as the likelihood evaluated at the maximal likelihood parameter set divided by the square root of the log-determinant of its local Hessian (ignoring constant factors that cancelled or did not scale with the number of data points [Bishop, 2016]). This is a more accurate approximation of the marginal likelihood than the often used Bayesian information criterion, as it uses information about the Hessian which was available in our case, see also above.

Within-participant object-consistency analysis: In order to test whether the correlation between performance on the two tasks across participants we found (Fig. 3.7, red) was not merely due to generic (e.g. attention-based) co-modulation effects, we performed a within-participant analysis of object-consistency. In particular, if correlation between performance in the two tasks is really driven by a unified underlying object representation, then the same pairs that participants regard as the true objects of the inventory during the visual familiarity test (and hence indicate as more familiar) should also be the ones that they treat as the true objects during the haptic pulling test (and hence pull harder in the direction parallel to their boundaries). Note that this reasoning is independent of the actual inventory that was set up in the experiment (Fig. 3.5A, inventory), and focuses on participants' internal representation, regardless whether it matched the actual inventory or not, only requiring that they behave consistently according to that internal representation in both tasks. In other words, this analysis is able to differentiate systematic deviations in participants' behaviour due to a misrepresentation of objects from

errors due to not having proper object-like representations.

To measure object-consistency within a participant, we calculated a haptic and a familiarity score for each unique scene that contained two true pairs in the haptic pulling test (12 and 4 in the visual and haptic exposure experiments, respectively), and computed the correlation between these scores across scenes. The haptic score was the average difference (across the repetitions of the same scene) in the pulling force participants generated when pulling to separate each of the two pairs into two compared to the pulling force generated to separate the two pairs from each other. The familiarity score was the average of the fraction of trials that the participant chose each of the pairs making up the scene as more familiar than another pair in the familiarity test. This score ranges from 0 (they never selected either pair in the familiarity test) to 1 (they always selected both pairs). We performed a t-test on these correlations across all participants, combining across experiments for statistical power. Participants who had a familiarity fraction correct of 1 (5 participants in the visual statistical exposure experiment) were excluded from this analysis as their object consistency-correlation was undefined.

Controlling for explicit knowledge of pairs: We also tested whether the generalization between visual and haptic statistics required explicit knowledge about the shape pairs. First, in order to quantify participants' explicit knowledge about the inventory, we computed the proportion of correctly identified true pairs (and ignored incorrectly identified pairs) based on the debriefing data (see section 3.3.2.1 Methods, Debriefing). As there were 6 true pairs in the visual, and 4 in the haptic statistical exposure experiment, the resulting scores were multiples of 1/6 or 1/4 for the two experiments, respectively (these were combined in Fig. 3.10). Next, we computed the correlation between their performance in the visual familiarity and in the haptic pulling test across the two experiments using multiple regression on visual performance with the two covariates being haptic performance and an indicator variable for the type of the exper-

iment (thus allowing for the average performances to depend on the experiment, but assuming that the regression slope was the same). Finally, partial correlations were measured between the performances in the two tests controlling for the proportion of correctly identified pairs (our measure of participants' explicit knowledge, see above). Partial correlation can reveal whether there is a significant relationship between the visual and haptic performance that cannot be explained by the explicit knowledge of the shape pairs. Specifically, in each experiment, both haptic and visual performance were regressed against explicitness. Residual performances in each modality were then computed by subtracting the performances predicted based on explicitness from the actual performances. The correlation between these residual performances across the two experiments was computed as for the raw performances and yielded our partial correlation measure. We also computed the ratio of the explained variances (R^2) of the normal and partial correlations in order to measure the extent to which the generalization effect could be explained by implicit transfer rather than by explicit knowledge.

Bayesian tests: In all statistical analyses we computed both the classical frequentist and the corresponding Bayesian tests. We used scaled JZS Bayes factors in the Bayesian t-tests, and in the Bayesian multiple linear regression for the correlational analyses with a scaling factor equal to $\frac{\sqrt{2}}{2}$ in the prior distribution (Morey & Rouder, 2011).

Exclusion criteria

In order to interpret the haptic pulling performance and its relation to visual familiarity performance (see above), it was essential that participants understood the general rules of pulling and scene breakage in our set-up (i.e. that objects were harder to break than to separate) which were used in all haptic task phases (haptic task training, haptic statistical exposure, and haptic pulling test). In the visual statistical exposure experiment, the only indicator of whether par-

ticipants understood the rules of pulling was their performance on haptic task training. Thus, in this experiment, participants were only included for further analysis if they had a significant (P<0.05) positive correlation between their pulling force and the required breakage force on clamp-catch trials of haptic task training. In contrast, in the haptic exposure experiment, pre-training with haptic task training only served to facilitate participants' learning in the subsequent haptic statistical exposure phase, in which they could also acquire an understanding of the rules of pulling, and so their haptic test performance itself was a reliable indicator of how much they understood these rules (as well as the identity of the pairs of the inventory). As we used the full range of haptic test performance to predict performance in the visual familiarity test (Fig. 3.7, red lines, see also below), not understanding the rules of pulling could not lead to an erroneous negative finding. Therefore there was no need to exclude any of the participants in this experiment based on their performance on haptic task training. Nevertheless, we repeated all analyses by excluding participants based on the same criteria as in the other experiment (leading to the exclusion of only one participant), and all our results remained essentially unchanged, with small numerical modifications to the test statistics (not shown).

3.3.2.2 Results

In line with previous results (Fiser & Aslin, 2001; Fiser & Aslin, 2005), we found that mere visual observation of the exposure scenes enabled participants to perform significantly above chance in the visual familiarity test (Fig. 3.7A, black dots: visual familiarity performance for individuals, green dot and error bars: group average quantified by fraction correct 0.77 [CI_{95}: 0.67-0.87], t_{19}=5.66, P=$1.9 \cdot 10^{-5}$, Bayes factor=1253). That is, in novel test scenes participants judged "true pairs" more familiar than "pseudo pairs", despite having seen all constituent shapes an equal number of times.

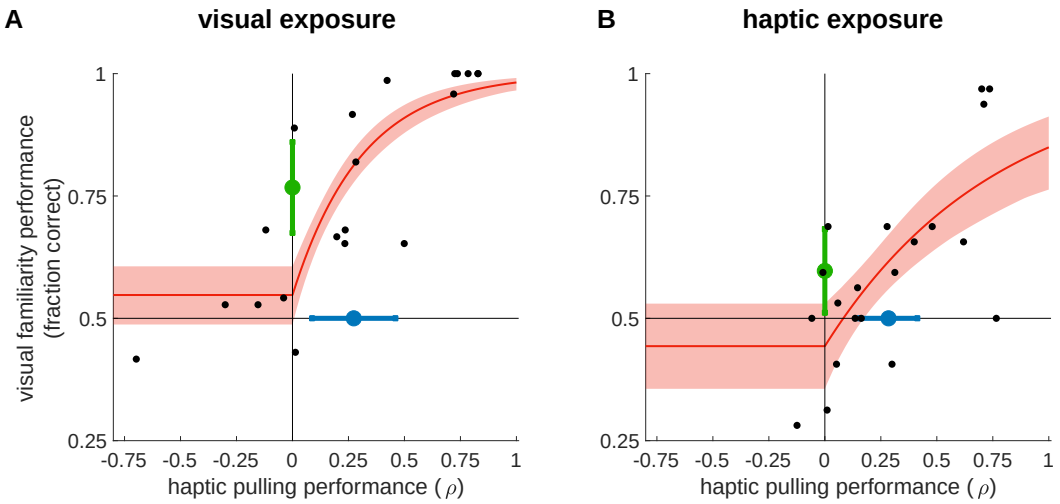


Figure 3.7: Learning from exposure to visual **(A**, Experiment 1**)** and haptic statistics **(B**, Experiment 2**)**. Performance on the visual familiarity test against haptic pulling performance for individual participants (black dots) with rectified exponential-binomial fit (red ± 95% confidence limits). Visual familiarity performance was measured by the fraction of correct responses (selecting true over pseudo pairs). Haptic pulling performance was quantified as the correlation coefficient (ρ) between the true breakage force and participants' pulling force across test scenes. Average performance (mean ± 95% confidence intervals) across participants in the two tasks is shown by coloured error bars (familiarity: green, pulling: blue). Vertical and horizontal lines show chance performance for visual familiarity and haptic pulling performance, respectively. Note that in the first experiment **(A)**, the performance of two participants was identically high in both tasks, and thus their data points overlap in the top right corner of the plot. Adapted with permission from G. Lengyel, G. Žalalytė, A. Pantelides, J. N. Ingram, J. Fiser, M. Lengyel and D. M. Wolpert.

On the haptic pulling task participants' performance was measured as the correlation (ρ) between their pulling force and the required breakage force (see Fig. 3.8). Participants performed significantly above chance (Fig. 3.7A, black dots: haptic pulling performance for individuals, blue dot and error bars: group average ρ=0.27 [CI_{95}: 0.07-0.47], t_{19}=2.88, P=0.0095, Bayes factor=5; see also Fig. 3.8A). While this effect was weak on average, more importantly, across participants, there was also a highly significant positive relationship between their performance on the visual familiarity and haptic pulling test (Fig. 3.7A, red, rectified exponential-binomial fit, likelihood ratio test, $\chi^2(2)$=265, $P<1 \cdot 10^{-10}$, $\log_{10}$Bayes factor=141). In particular, our fit of the data revealed that going from random haptic pulling performance (ρ=0) to perfect performance (ρ=1) covered 87% of the possible range of visual familiarity performance

(fraction correct from 0.5 to 1.0). We also tested whether there was a significant change in performance throughout the haptic pulling test trials and found no significant trend (P=0.07, Bayes factor=1) suggesting that generalization immediately appeared and it did not just gradually develop during the test. These results show that participants who learned in one modality successfully generalised what they learned through visual statistics to predict the haptic properties of objects, and suggest that variability in performance on both tasks across participants is due to the same underlying cause: differences in how well participants learned the inventory.

3.3.3 Experiment 2 - Haptic-to-visual generalization

In the second experiment, we examined haptic learning and haptic-to-visual generalisation with a different group of N=20 participants (Fig. 3.5A, inventory, bottom). As in the previous experiment, in order to familiarise participants with our setup, we pre-trained them on the basic pulling paradigm with coloured rectangles as objects, without any reference to the shapes of the inventory (Fig. 3.9). They were then exposed to the haptic statistics of the inventory (Fig. 3.5A, exposure, bottom; Fig. 3.5B, left). During exposure, each scene consisted of 4 shapes arranged as a 2×2 block on a grey square (Fig. 3.5A, exposure, bottom), and participants were required to pull apart these scenes in predefined directions so as to part the scene into two equal pieces (Fig. 3.5B, left; i.e. into two vertical pieces for horizontal pulling and vice versa). Again, unknown to the participants, each scene was constructed by combining two of the objects from the inventory (either a pseudo-pseudo, a pseudo-true, or a true-true combination of pairs arranged either vertically or horizontally). We chose to have only 2 and not 3 objects in each scene so that participants always knew the scene would break apart simply into two pieces — the physics of multiple objects with complicated (potentially non-convex) geometries would have been much more difficult to simulate and expect participants to understand.

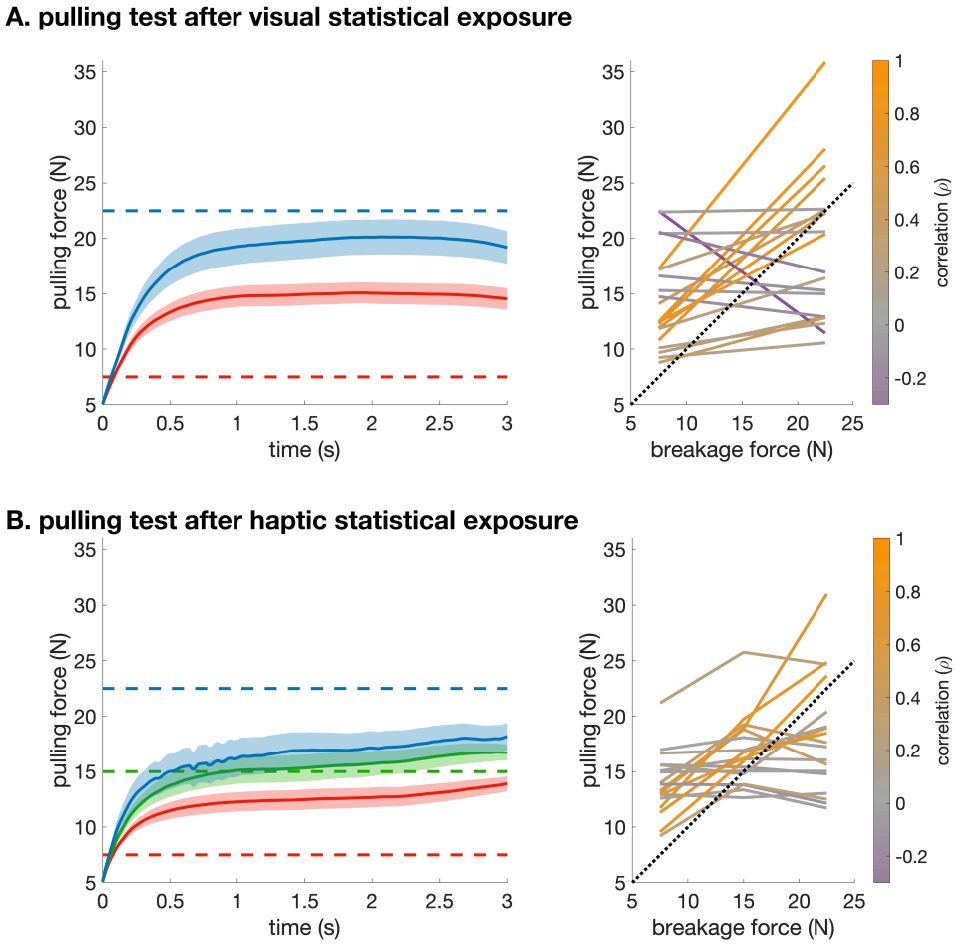


Figure 3.8: Pulling performance in the visual **(A**, Experiment 1**)** and haptic **(B**, Experiment 2**)** statistical exposure experiment. **(A)** Left: Force traces from the start of pulling (5 N) on clamp-catch trials in the haptic pulling test of the visual statistical exposure experiment. Data shows mean±s.e.m. across participants for trials in which the breakage force was high (blue) or low (red). Dashed lines show the corresponding breakage forces. Right: Average pulling force (at 3 s) vs. breakage force (2 levels) for each participant colour coded by their correlation (across all trials). Dotted line shows identity. The average pulling force difference between the two levels was 9.3 N ± 3.3 N (s.e.m.). **(B)** as A for the clamp trials in the haptic pulling test of the haptic statistical exposure experiment, in which there were three levels of breakage force. The average pulling force difference between the low and medium breakage force levels was 5.4 N ± 1.2 N (s.e.m.), and between the medium and high breakage force levels was 2.8 N ± 1.7 N (s.e.m.). Raw data necessary to generate this figure was only saved for 18 participants. Adapted with permission from G. Lengyel, G. Žalalytė, A. Pantelides, J. N. Ingram, J. Fiser, M. Lengyel and D. M. Wolpert.

Critically, the force (simulated by the bimanual robotic interface) at which each scene separated depended both on the constituent pairs and the pulling direction, and only true pairs behaved haptically as unitary objects in that their shapes required more force to separate than

the shapes of pseudo pairs, or shapes belonging to different pairs. This led to three different force levels required to separate the scenes, with the lowest force when pulling apart any combination of pairs orthogonal to their boundary (Fig. 3.5A, Haptic exposure: examples 1 and 3, 7.5 N), the highest force required when separating two true pairs into their constituent pieces (2nd example, 22.5 N), and an intermediate force when separating a true and a pseudo pair into their constituent pieces (4th example: 15 N). As participants pulled on each side of the scene, the resistive force generated by the robots rose until it reached a threshold (7.5, 15 or 22.5 N depending on the scene), at which point the forces dropped to zero and the scene parted visually. The shapes were masked from just before pulling started until the scene was successfully separated. Thus, the duration for which the shapes were seen as unseparated and then separated also conveyed no information about the identity of the true and pseudo pairs. Importantly, although these participants had visual experience with the objects, true and pseudo pairs appeared the same number of times and with the same consistency (i.e. their constituent shapes always appeared together in the same spatial configuration), and so visual information could not be used to distinguish between them. Therefore, objects (true pairs) could only be identified by the physical effort required to pull the scenes apart.

Following the exposure, participants were tested on the same two tasks as in the previous experiment: (1) the haptic pulling task measuring within-modality learning and (2) the visual familiarity test assessing across-modality generalization. Participants successfully learnt the haptic exposure and built representations about which scenes required more or less pulling force. Moreover, the haptic exposure influenced participants' judgements in the visual familiarity test demonstrating that the units/chunks learnt during the haptic pulling task generalized to a purely visual discrimination task.

3.3.3.1 Methods

Participants

20 participants (age range 21-34, mean 26, 16 women) gave informed consent and participated. No participants were excluded. Data was collected in two installments. First, we made a preliminary estimate of the approximate number of participants we would need for significant results and collected data accordingly. This resulted in 13 participants. In the same way as in Exp. 1, additional participants (7 more in exp. 2) were recruited after the suggestion of an external expert. All our main results remained highly significant. Again, the process of adding participants did not biased our results due to the consistent usage of Bayes factors (Dienes, 2011). All experimental protocols were approved by the University of Cambridge Psychology Ethics Committee.

Haptic statistical exposure experiment

The experiment consisted of four phases: (1) haptic task training, (2) haptic statistical exposure, (3) haptic pulling test, and (4) visual familiarity test. Note that the ordering of the main phases of the experiment (statistical exposure $\rightarrow$ haptic testing $\rightarrow$ visual testing) remained identical across the two experiments (Fig. 3.5A). However, the ordering of the haptic task training phase was chosen so that it immediately preceded that phase of the experiment in which haptic experience was first combined with the shapes of the inventory, i.e. the haptic statistical exposure phase in this experiment and the haptic pulling test in the visual statistical exposure experiment.

Haptic task training: This was similar to the haptic task training in the visual statistical exposure experiment, except that scenes could include not only two differently coloured 2-by-1 rectangles as before (C2: i.e. 2 colours) but also one 2-by-1 rectangle and two 1-by-1 squares (C3), or four 1-by-1 squares (C4, Fig. 3.9). All these configurations were arranged

in a 2-by-2 block of coloured pieces as before, and as all "objects" (rectangles or squares) were differently coloured they still had clear, visually identifiable boundaries between them (4 colours used in total). The additional configurations were needed as the haptic statistical exposure included pseudo as well as true pairs, where pseudo pairs behaved haptically as two separate single elements, and so three rather than two force levels were possible [see below] which thus needed to be demonstrated during haptic task training. The required minimal pulling forces were determined as above. This meant that the same two force levels (7.5 and 22.5 N) were needed to pull apart C2 scenes in the easy (orthogonal to the boundary between the rectangles) and hard directions as in the visual statistical exposure experiment (see above), while C4 scenes were easy (7.5 N) to pull apart in either direction, and C3 scenes were easy (7.5 N) to pull apart in the direction orthogonal to the long side of the rectangle and medium hard (15 N) in the other direction.

Participants completed a total of 144 trials, which consisted of 96 Standard trials (composed of two 48-trial blocks), followed by 48 Clamp trials. (Clamp-catch trials were omitted as it was not necessary to include them in the haptic pulling test, see below.) Trials within each block consisted of 8 trials with C2, 8 trials with C4 and 32 trials with C3 in a pseudorandom order and orientation. These proportions were chosen to match those used in the haptic statistical exposure phase, see below.

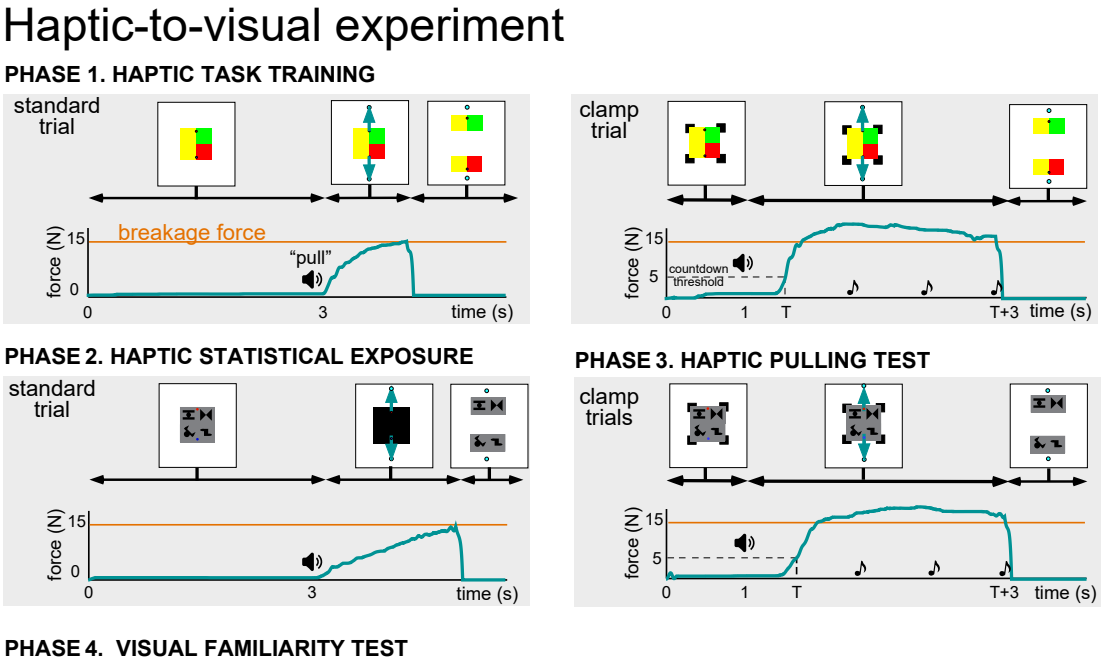


Figure 3.9: Phases of the haptic statistical exposure experiment. Phase 1. Haptic task training consisted of different combinations of coloured blocks (2-4; see section 3.3.3.1, Methods) and the breakage force (3 levels) depended on the configuration. Standard and clamp trials were the same as in the visual statistical exposure experiment (Fig. 3.6). **Phase 2.** Haptic statistical exposure consisted of standard trials with scenes consisting of the true and pseudo pairs of the inventory. Just prior to the initiation of pulling, the scene was masked and only unmasked when the pulling force exceeded the breakage threshold and the scene separated. **Phase 3.** Clamp trials in the haptic pulling test phase also used scenes consisting of true and pseudo pairs of the inventory. The scene was only masked after the clamps were removed if the force was insufficient to separate the scene. The mask was removed once the scene was separated. The use of masks in haptic statistical exposure and pulling test ensured that the time the scene was seen both together or apart was independent of the breakage force. **Phase 4.** The visual familiarity test consisted of 32 trials (not shown, see Fig. 3.5A) and was as described before in Fig. 3.6. Adapted with permission from G. Lengyel, G. Žalalytė, A. Pantelides, J. N. Ingram, J. Fiser, M. Lengyel and D. M. Wolpert.

Haptic statistical exposure: This phase was similar to the haptic test in the visual statistical exposure experiment but included both true and pseudo pairs, and only Standard trials. Specifically, each visual scene could be composed of either two true pairs, or a true pair and a pseudo pair, or two pseudo pairs, such that the two pairs always had the same orientation, touching on the long side, thus forming a 2×2 block of four shapes without a visible boundary between the pairs. Critically, pseudo pairs were indistinguishable from true pairs based on their

visual appearance statistics: they appeared the same number of times, in the same combinations with the other pairs. This was important so that any consistent preference in the visual familiarity test (see below) between true and pseudo pairs could have only been due to their different haptic statistical properties. Specifically, pseudo pairs behaved haptically as two separate single-shape objects, rather than one integrated object, so that the constituent shapes of a pseudo pair were as easy to pull apart as shapes belonging to two different objects. This meant that three force levels were required: two true pairs were hard (22.5 N) to pull apart in the direction parallel to their boundary and easy (7.5 N) in the other direction, two pseudo pairs were easy (7.5 N) to pull apart in either direction, and a true and a pseudo pair was easy (7.5 N) to pull apart in the direction orthogonal to the long side of the two pair and medium hard (15 N) in the other direction.

In order to ensure that the time for which each scene was presented in an unseparated and separated state was independent of how much time participants spent on pulling it apart, in each trial, the 2×2 block of four shapes was masked 3s after the hands were moved into their home positions (i.e. just before pulling could start) and unmasked once the scene was successfully separated. Note that according to the rules of the task (see above) all trials ended by the scene eventually becoming separated, regardless of its composition and the pulling direction. Thus, the visual statistics of the scenes remained independent from their haptic properties and conveyed no information about the identity of the true and pseudo pairs. The instructions simply told participants that "the force required to break the block apart in each direction will depend only on the symbols and their configuration on the block" and asked them to "learn the minimal force required to pull the block apart in each direction and we will test you on this later". (Note that, in contrast to other phases of the experiment involving haptic manipulations, no Clamp or Clamp-catch trials were needed in this phase as we were only exposing participants to haptic

statistics but not yet measuring their performance — which occured in the next phase the haptic pulling test, see below.)

Participants completed 192 Standard trials (composed of four 48 trial blocks). Each block of trials included all possible combinations of two pairs of the same orientation in both pulling directions. This meant that scenes with two true pairs were presented 8 times (4×22.5 N trials and 4×7.5 N trials), scenes with two pseudo pairs were presented 8 times (8×7.5 N trials) and scenes with a true and a pseudo pair were presented 32 times (16×7.5 N trials, 16×15 N trials). Trials within each block were randomized.

Note that there were fewer trial scenes in the haptic than in the visual statistical exposure experiment because less unique scenes could be generated in the 2×2 arrangement. Moreover, due to the time the robotic interface needed to shift from one pulling position to the other, the presentation time of the scenes was longer in the haptic exposure than in the visual statistical exposure experiment (Fig. 3.9).

Haptic pulling test: In order to measure how much participants learned from haptic statistical exposure, we tested their haptic performance as in the other experiment. Therefore, this phase was similar to the haptic pulling test in the visual statistical exposure experiment but included both true and pseudo pairs as did the haptic statistical exposure phase of this experiment, and used Clamp-trials rather than Clamp-catch trials. (Clamp-catch was unnecessary here as there was no need to prevent participants gaining additional haptic information from these trials in this experiment.) Again, to ensure that each scene could be seen in an unseparated and separated state for a fixed amount of time, irrespective of its haptic properties, it was masked during the period between the removal of the clamps and the separation of the scene.

Participants completed one block of 48 Clamp trials which were similar to one block of Standard trials in the haptic statistical exposure phase, except for the presence of the clamps.

Visual familiarity test: These trials were identical to the familiarity test in the visual statistical exposure experiment. Participants completed 32 trials (2×16-trial blocks), such that within every 16-trial block each true pair was compared to each pseudo pair of the same orientation in each order exactly once (the order of trials was randomly permuted within each block). Again, this last phase of the experiment was the first time participants saw individual, separated objects (constructed from the shapes of the inventory), of which the boundary was thus obvious.

Although the assignment of shapes to objects (pairs) was randomized across participants, we found at the end of the experiments that some participants had the same order of trials due to a coding error. Specifically, in the haptic statistical exposure experiment, two participants shared the same haptic exposure sequence. In the visual statistical exposure experiment, three participants shared the same haptic testing sequence and two shared the same visual familiarity testing sequence. There is no reason to believe that the order of trials would affect learning or performance.

Data analysis

We performed the same data analysis as in Experiment 1. See section 3.3.2.1 Data analysis, Methods, Experiment 1.

3.3.3.2 Results

Performance on the haptic pulling test showed that participants successfully learned which scenes required more or less pulling force (Fig. 3.7B, black: haptic pulling performance for individuals, blue: group average ρ=0.28 [CI_{95}: 0.14-0.42], t_{19}=4.34, P=$3.8 \cdot 10^{-3}$, Bayes factor=91; see also Fig. 3.8B). Haptic experience also affected participants' judgements in the visual familiarity test, in which they needed to compare two pairs, one a true pair and the other a pseudo pair. Participants judged true pairs significantly more familiar than pseudo pairs (Fig. 3.7B,

black: visual familiarity performance for individuals, green: group average quantified by fraction correct 0.6 [CI_{95}: 0.51-0.69], t_{19}=2.2, P=0.038). Note that this across-modality effect was even weaker on average than previously (Bayes factor=2 indicates evidence that is weak or not worth mentioning) because, in contrast to the previous experiment, haptic and visual statistics were now in explicit conflict: true and pseudo pairs (compared in the visual familiarity task) were identical in their visual statistics and only differed in their haptic statistics. As there was no haptic stimulus during visual statistical exposure in the other experiment, no such conflict arose there.

More critically, we also found again that participants' familiarity performance had a highly significant positive relationship with their pulling performance (Fig. 3.7B, red, rectified exponential-binomial fit, likelihood ratio test, $\chi^2(2)$=47.2, P=$5.6 \cdot 10^{-11}$, $\log_{10}$Bayes factor=35), such that performance on the haptic pulling test accounted for 81% of visual familiarity performance. As before, there was no significant change in performance throughout the familiarity test trials (P=0.58, Bayes factor<1) suggesting that the generalization effect did not gradually emerge during the test trials. These results parallel the results of the visual exposure experiment. Moreover, they demonstrate a particularly strong form of generalisation of information acquired through haptic statistics to judging visual properties of objects — at least in those participants who learned the haptic statistics well. That is, objects that appeared precisely the same number of times as others were "illusorily" but systematically perceived as visually more familiar just because they had more object-like haptic properties. Interestingly, we found similar levels of haptic performance in the two experiments (t_{18}=0.09, P=0.93, Bayes factor (favoring the same performance levels)=3) even though in the first experiment there was no haptic statistical exposure at all and participants' haptic performance relied only on generalization from the visual exposure. Performance on the visual familiarity test was higher after visual expos-

ure than after haptic exposure (t_{18}=2.65, P=0.01, Bayes factor (favoring different performance levels)=4) which was expected based on the fundamental difference in cue conflicts between the two experiments.

In order to test whether the positive relationship between performance on the two tasks across participants (Fig. 3.7A and B, red) was not merely due to generic (e.g. attention-based) sources of modulation, we performed a within-participant analysis of object-consistency (see section 3.3.2.1, Methods). This analysis measured, for each participant, how much the particular pairs they regarded as the true objects of the inventory during the visual familiarity test (and hence indicated as more familiar) were also the ones that they treated as the true objects during the haptic pulling test (and hence pulled harder when needed to break them). This was quantified by a single scalar measure (correlation) between familiarity and pulling force for individual scenes as a measure of consistency. As this was a noisy measure, based on a limited number of trials with each participant, we then pooled the data from both experiments and used a t-test across the participants to ask if this measure was significantly different from zero. We found a significantly positive consistency (correlation R=0.297 ± 0.104 with t_{34}=2.86, P=0.007, Bayes factor=6). Taken together, this demonstrates that participants developed a modality-generic representation of objects from either visual or haptic statistical contingencies alone, which in turn they could transfer to the other modality.

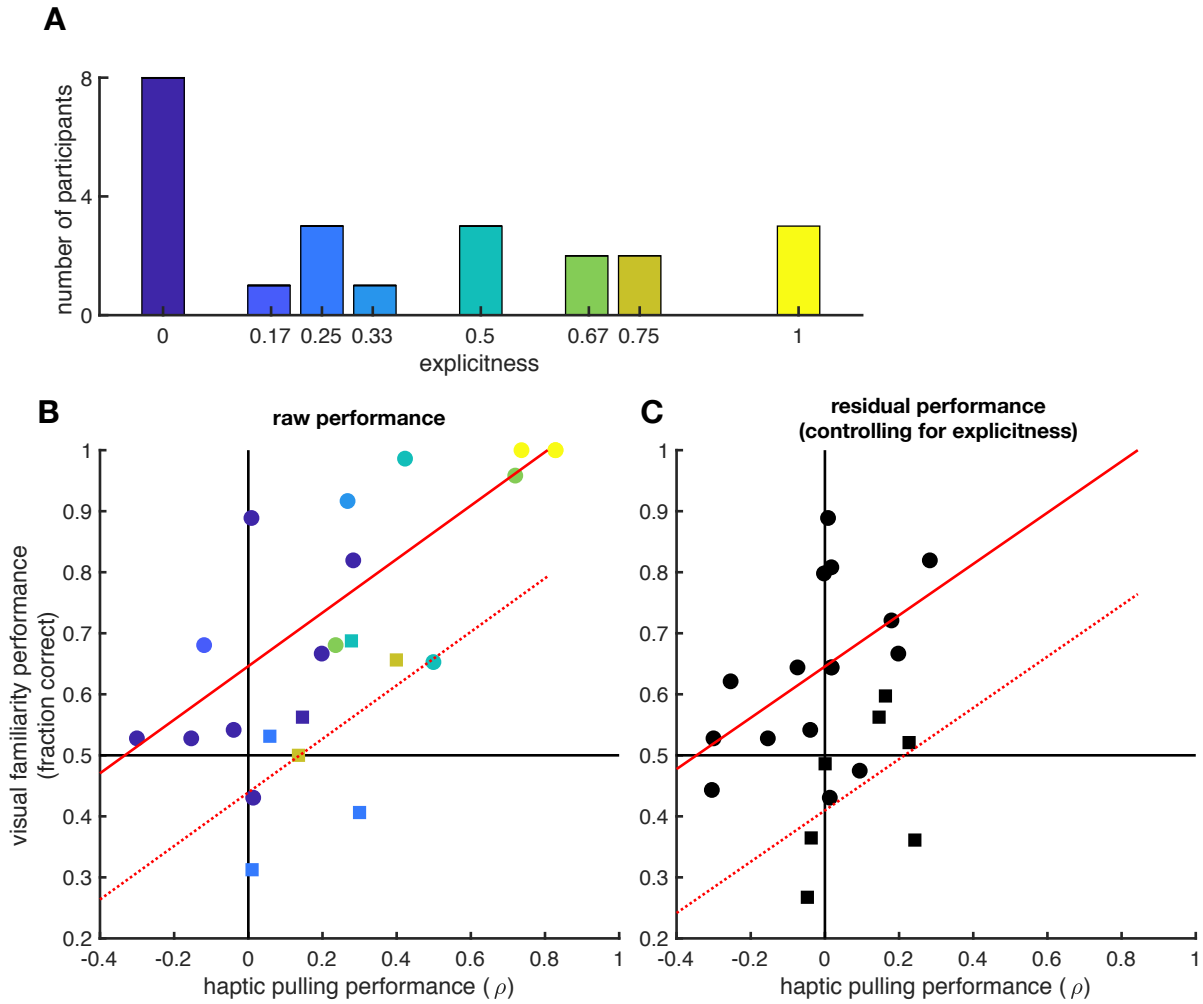


Figure 3.10: Effects of explicit knowledge on generalisation. Participants' explicit sense of knowledge was quantified as the proportion of true pairs they correctly identified (out of 6 in the visual statistical exposure and out of 4 in the haptic statistical exposure experiment, resulting in 9 possible unique levels in total, out of which 8 were realized) during a debriefing session following the experiment (N=23 participants). **(A)** Histogram of explicitness across participants (average=0.37). **(B)** Visual and haptic performance as in Figure 2, pooled across the two experiments for those participants who were debriefed (circles: visual statistical exposure, squares: haptic statistical exposure experiment). Colors show explicitness for each participant as in panel A. Red lines show linear regression assuming same slope but allowing for different average performances in the two experiments (solid: visual statistical exposure, dotted: haptic statistical exposure experiment): R=0.84 (CI_{95}: 0.65-0.93), P=$6.2 \cdot 10^{-7}$. **(C)** Residual visual and haptic performance after controlling for explicitness (symbols as in panel B). In each experiment, both haptic and visual performance were regressed against explicitness. Residual performances in each modality were then computed by subtracting the performances predicted based on explicitness from the actual performances. Red lines show linear regression as in panel B: R=0.69 (CI_{95}: 0.38-0.86), P=$3.1 \cdot 10^{-4}$. Adapted with permission from G. Lengyel, G. Žalalytė, A. Pantelides, J. N. Ingram, J. Fiser, M. Lengyel and D. M. Wolpert.

Finally, we tested whether the generalization between visual and haptic statistics required an

explicit sense of knowledge about the shape pairs (Fig. 3.10). Quantitative debriefing data were collected from 23 participants after the experiments from which we computed the proportion of correctly identified true pairs as a measure of explicit knowledge (Fig. 3.10A). Across these participants, the performance in the visual familiarity and in the haptic pulling test strongly correlated (R=0.84 [CI_{95}: 0.65-0.93], P=$6.2{\cdot}10^{-7}$, Bayes factor=3225, see also Fig. 3.10B). Critically, when we controlled for participants' explicit knowledge (proportion of correctly identified pairs) on the relationship between visual and haptic performance, we still found a highly significant partial correlation (R=0.69 [CI_{95}: 0.38-0.86], P=$3.1 \cdot 10^{-4}$, Bayes factor=23.4, see also Fig. 3.10C) suggesting strong implicit transfer between modalities in addition to that afforded by this kind of explicit knowledge. Furthermore, the ratio of the explained variances (R2) shows that the larger part (67%) of the generalization effect is due to implicit transfer and cannot be explained by explicit reasoning about the pairs.

3.3.3.3 Discussion

In summary, we found evidence that participants could segment scenes into objects based on either visual or haptic statistics alone, without any boundaries that could identify the objects. Such learning led to genuinely coherent object-like representations as participants segmented scenes into objects consistently across the two modalities, independent of the modality in which the statistics of the objects were originally experienced. Our participants' within- and across-modality performance was not perfect as implicit statistical learning over short periods is known to be difficult (Kim et al., 2009; Perruchet & Pacton, 2006). However, critically, participants who learned well within one modality showed strong generalisation to the other modality (Fig. 3.7), beyond what an explicit sense of knowledge of the objects, potentially leading to highly cognitive strategies, would have predicted (Fig. 3.10).

Earlier reports in statistical learning only showed that statistical cues can be used for segmentation (in various sensory modalities). However, they typically focused on a single modality (Conway & Christiansen, 2005; Creel et al., 2004; Fiser & Aslin, 2001; Hunt & Aslin, 2001) or showed that humans can combine statistical information simultaneously presented in more than one modality (Conway & Christiansen, 2006). Critically, these studies did not investigate the "objectness" of the resulting representations in any way. In particular, they did not test generalisation across modalities and hence could not exclude the possibility that performance in each modality only relied on information presented in that modality alone, without an underlying modality-general object-like representation. Conversely, other studies showed generalization across visual and haptic modalities, but they used objects which were already fully segmented by low-level boundary cues and as such they could not investigate the role of statistical learning in the emergence of object-like representations (Yildirim & Jacobs, 2013). Instead, our findings suggest a deeper underlying integration of object-like representations obtained by statistical learning: any statistically defined structural information obtained in one modality becomes automatically integrated into a general internal representation linking multiple modalities.

Although our experiments were conducted with adult participants, infants have also been shown to learn to segment visual scenes or auditory streams automatically, after mere passive exposure (Fiser & Aslin, 2002; Kirkham et al., 2002; Quinn & Bhatt, 2005; Saffran & Kirkham, 2018). Importantly, these studies used stimuli with a statistical structure (and in the case of visual experiments, actual constituent shapes) that were similar or even identical to those used in our experiment (Fiser & Aslin, 2002; Kirkham et al., 2002; Quinn & Bhatt, 2005; Saffran et al., 1996; Saffran & Kirkham, 2018). This suggests that, by parsimony, infants possess the same sensitivity to the co-occurrence statistics of sensory inputs as adults (Aslin, 2017). Since we showed that statistical learning produces object-like representations in adults, we propose

that the statistical learning mechanisms revealed in our experiments might also operate in the emergence of object representations during cognitive development.

If, as we argue, the statistical learning mechanisms we revealed also operate in infants, the present findings complement the results of earlier infant studies on object representations. We tested whether humans can use primarily statistical cues to segment the world into constituent components which pass a fundamental criterion for objects — that of zero-shot across-modality generalization (i.e. going beyond observed statistical regularities; Spelke (1990)). In contrast, previous studies of cognitive development defined a specific set of criteria, including cohesion, boundedness, rigidity, and no action at a distance, that infants use to identify objects (Kellman & Spelke, 1983; Spelke, 1990). Our results suggest that these criteria may be sufficient but not necessary. For example, one might argue that the objects in our experiment violated even the basic requirement of having 3-dimensional structure, and specifically the principle of "cohesion" of Spelke (1990) because their constituent shapes were separated by gaps (although in front of a spatially contiguous gray background). Thus, these classical criteria may be special cases of a more general principle of statistical coherence. Nevertheless, an internal object-like representation segmented based on statistical coherence (and other cues) may need to eventually pass a number of additional criteria (e.g. those involving cohesion) to become a real mental object, and it will be for future studies to test whether and how statistical learning mechanisms can produce such representations.

In general, there may be many shades of perceiving "objectness" (ranging from rigid bodies through more elusive entities, such as a melting scoop of ice cream or the jet of steam of a boiling teapot, to collections of clearly separate objects). Thus, further work will be needed to refine the necessary and sufficient conditions for segmenting entities with different degrees of objectness on this continuum. Similarly, the difference in the behavioral measures used to

index object perception in previous studies and in our experiments (looking times vs. across-modality generalisation) may also need more attention. Specifically, it will be interesting to see whether the perception of (a degree of) objectness is always reflected consistently in all forms of behavior, or it is subject to paradoxical effects, akin to e.g. the size-weight illusion (Flanagan & Beltzner, 2000), when the same feature seems to be perceived differently in the context of controlling different aspects of behavior (decision making vs. motor control). From this perspective, our work represents an important first step in connecting the field of statistical learning to the kind of object representations that have been identified in infants.

Finally, although the present study does not provide empirical evidence for a single specific cognitive mechanism underlying the generalization effects we found, these results together with previous studies (Lake et al., 2015; Orbán et al., 2008) point to possible computations explaining the present findings. First, the generalisation effects occurred without any ancillary cues that are required to engage specialized learning mechanisms, such as segmentation cues for implicit rule learning (Peña et al., 2002; Saffran et al., 2007), verbal instructions for explicit hypothesis testing (Shanks, 2010), or ostensive signals for social learning (Csibra & Gergely, 2006). Second, it is unlikely that participants were able to retain in memory all the raw sensory stimuli they received during the exposure phase (e.g., 444 scenes with 6 shapes in each for the statistical exposure experiment). Thus, they must have developed some compressed representation of those stimuli during exposure, and it is only this representation that then could allow them to generalise in the test phase. Third, with regard to the form of the compressed representation, statistical learning goes beyond the learning of simple (pairwise) associations between the constituent components of objects, and has been shown to be best described as the extraction of statistically meaningful (potentially multivariate) latent "chunks" (Gershman & Niv, 2010; Orbán et al., 2008; Yildirim & Jacobs, 2012). Therefore, we propose that these latent

chunks are the abstract representations that are built automatically during exposure and mediate the across-modality effects we observed. Accumulating evidence supports this view by showing that the neural representation underlying multimodal integration might involve cortical areas traditionally linked to unimodal processing (Amedi et al., 2001; Ghazanfar & Schroeder, 2006). Together, these results suggest that statistical learning is not only a domain-general mechanism (Frost et al., 2015; Kirkham et al., 2002; Thiessen, 2011), but it also results in domain-general internal representations that could be the basis for the emergence of affordances (Gibson, 1979; Parker & Gibson, 1977) and the abstraction of object concepts (Carey, 2009b; Leslie et al., 1998; Spelke, 1990).

3.4 Conclusion

The two studies in this chapter demonstrate that in classical SL paradigms participants create abstract, amodal representations of chunks that serve as perceptual units for processing subsequent sensory input in a similar way to how the representations of real objects form units for perceptual processing. In the first study, I showed that the representations of statistically defined chunks learnt during SL elicited similar attentional and perceptual processes to what real objects elicited. In the second study, I presented empirical evidence showing that the representation of statistically defined chunks were abstract enough to be amodal that allowed zero-shot across-modality generalization.

Combining these results with the findings of previous studies showing that the sensitivity to most sensory cues responsible for object segmentation develops later during infancy (Spelke, 1990) and that the statistical learning mechanisms we revealed here also operate in infants (Fiser & Aslin, 2002; Kirkham et al., 2002; Saffran et al., 1996) leads to the parsimonious proposal that learning the consistent statistical properties in the environment has a key role in

the emergence of object representations during infancy. The relative importance of statistically learned and innately available cues and representations in the development of real mental objects remains to be investigated in future studies.

The proposal, that SL has an important role in mental object formation, also suggests that we need to reconsider the definition of "objectness". If objects are defined as a sufficient set of statistical contingencies, then objectness will be defined on a continuous scale and the degree of objectness will depend on the level of statistical coherence. In most natural settings, objectness rather seems to be an all-or-none feature without uncertainty, but there are several real life example when objectness is indeed ambiguous, e.g. think about a jet of steam, the illusory contours of the Kanizsa triangle, or any example of animal camouflage. Thus, results of the two studies in his chapter encourage the field of developmental cognitive science to refine the necessary and sufficient conditions for segmenting objects that might have different degrees of objectness in the eye of an infant.

Finally, regarding the mechanism that are involved in building object-like representations of statistically defined chunks, we argue in both studies that since a probabilistic chunking mechanism could capture the underlying computations the best in previous studies using very similar paradigms (Orbán et al., 2008) the learning mechanism involved in the present studies could be described by the extended version of the probabilistic chunk learning mechanism proposed in Orbán et al. (2008). Being a Bayesian latent variable model, the computational model in Orbán et al. (2008) can be directly extended to the HBM framework, proposed for perceptual and statistical learning in Chapter 1.

After investigating classical PL (in Chapter 2) and SL (in Chapter 3) paradigms the next chapter focuses on the interaction between two domains of learning using a PL paradigm called roving. I will demonstrate that several previously unexplained phenomena can parsimoniously

be explained by assuming that the SL process interacts with the PL process both of which can be captured jointly under the same HBM.

Chapter 4

Bayesian Statistical Perceptual Learning

4.1 Summary

Roving in PL refers to paradigms, in which the properties of the perceptual attribute in focus are intermixed during training. Most previous investigations studied paradigms where multiple reference-stimuli were intermixed in a discrimination task and found that PL were disrupted when the references were interleaved on a trial-by-trial basis, while PL were intact when the references were separated into blocks or were interleaved across the trials but the mixing followed a fixed temporal order. In this chapter, I will account for this pattern of results by assuming that the observer also learns the temporal structure of the reference sequence via SL and this knowledge then interacts with and supports PL. Following the framework suggested in Chapter 1, I will treat PL and SL under the same hierarchical Bayesian model (HBM) and formalize a Bayesian Statistical Perceptual Learning (BSPL) model that can accommodate classical and roving studies in PL. The BSPL model is based on the contextual inference model in Heald et al. (2020) and assumes a hidden Markov SL learning the transition model of the reference-contexts and a PL process optimizing neural resource allocation that modulates the stimulus

encoding to maximize discrimination performance.

As a first step, I will use simulated data and demonstrate that the BSPL model can capture the wide range of behaviour results reported in roving experiments. I found that inferring the reference-context of the trials and learning the transition model between the different reference-contexts could substantially support PL in the blocked and the fixed order roving conditions, while such context learning did not have a positive effect in the randomly interleaved condition. Based on these results, I suggest that the new HBM framework can capture most of the previously unexplained phenomena in PL obtained by using more complex stimuli. Furthermore, since naturalistic learning scenarios always involve both learning some relevant structures and adapting perception to those learned structures, the BSPL framework jointly capturing both PL and SL provides a parsimonious computation approach for sensory learning.

4.2 Roving in perceptual learning

In PL paradigms, roving refers to conditions in which some properties of the task are intermixed during training (Dosher et al., 2020; Zhang et al., 2008). In this chapter, I will focus on the most widely used roving paradigm - discrimination task with multiple references. In the classical discrimination tasks of PL, participants are trained to discriminate the perceptual feature values of the test stimuli from the feature value of the reference stimulus (see Fig. 1.1A). In this classical setup, the feature value of the reference stimulus is fixed during the entire training, while the test stimulus can have many different feature values ranging from very close to far away compared to the value of the reference. In contrast, roving experiments have multiple different reference values and participants are trained to discriminate the test-stimuli from all these references. For example, in the orientation discrimination task shown in Fig. 4.1A, the reference orientation in each trial is chosen from four different reference angles during the training.

Roving conditions in discrimination tasks with multiple references can be grouped broadly into three categories: blocked, randomly interleaved, and fixed order (see Fig. 4.1B). In the blocked condition, the discrimination trials with the same reference values are grouped together into the same block separately from the other blocks (Fig. 4.1B1). Therefore, the observers practice in one block of trials with a single reference, then switch to another block of trials with another reference and so on. In the randomly interleaved condition, the reference value in a trial is randomly generated from a set of predefined references and thus, the reference in the discrimination task changes in each trial during the practice (Fig. 4.1B2). Finally, in the fixed order condition, the reference changes in each trial, similar to the randomly interleaved condition, but the changes of the reference value across the trials follows a fixed order (Fig. 4.1B3). In the next section, I provide a brief summary of the results found in previous studies using these roving conditions and list the consistent pattern of results across the perceptual attributes and modalities that the present Bayesian Statistical Perceptual Learning (BSPL) model will address.

4.2.1 The pattern of results in roving paradigms

The diverse set of results across roving paradigms using multiple references points towards a consistent pattern that can be summarized as follows. There is no learning or the amount of learning is reduced when the references are randomly interleaved across trials during training. However, when the trials are grouped in blocks by reference values or the reference value changes trial-by-trial but follows a fixed order during the practice, PL emerges in the discrimination tasks. The BSPL model described in the next sections can account for this pattern of results. To set the stage for my modeling, I briefly review below the key results reported in the roving literature based on discrimination tasks with multiple references.

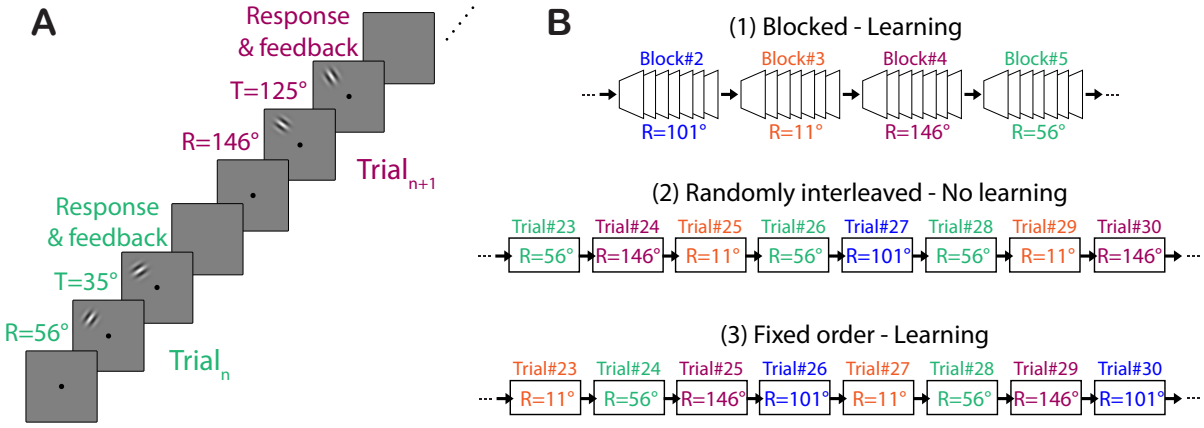


Figure 4.1: The paradigm and the results in roving conditions. Roving refers to perceptual discrimination tasks with multiple reference values. **A:** Two example trials in roving using an orientation discrimination task. R & T denote the angle of the reference- and the test-stimuli in degrees, respectively. The value of the reference changes across the trials, e.g., in trial n the reference is 56° but in the next trial, $n+1$, it changes to 146°. **B:** The three dominant conditions in roving paradigms: (1) when the references appears in blocks, PL emerges (Dosher et al., 2020; Nahum et al., 2010; Yu et al., 2004; Zhang et al., 2008), (2) when the references are randomly interleaved, PL dissapears (Adini et al., 2002; Adini et al., 2004; Amitay et al., 2005; Banai et al., 2010; Cong & Zhang, 2014; Dosher et al., 2020; Kuai et al., 2005; Nahum et al., 2010; Otto et al., 2006; Parkosadze et al., 2008; Tartaglia et al., 2009b; Yu et al., 2004; Zhang et al., 2008), and (3) when the references follow a fixed order, PL reemerges (Cong & Zhang, 2014; Kuai et al., 2005; Zhang et al., 2008)

During a debate about the conditions under which learning in contrast discrimination tasks might emerge (Adini et al., 2002; Adini et al., 2004; Yu et al., 2004), researchers found evidence for no or disrupted learning in contrast discrimination tasks with multiple reference values (Yu et al., 2004). Several follow-up studies replicated this lack of learning in discrimination tasks with multiple references using motion direction (Kuai et al., 2005), orientation (Dosher et al., 2020; Zhang et al., 2008), tone frequency (Amitay et al., 2005), temporal interval (Banai et al., 2010), line bisection (Otto et al., 2006; Parkosadze et al., 2008; Tartaglia et al., 2009b), and pseudo word identity (Nahum et al., 2010) as perceptual attributes. These results established the generality of the finding, and research in the field of PL started to focus on investigating this paradigm in more detail.

Experiments across multiple paradigms have demonstrated that PL emerged when the different references were grouped together into blocks during the practice (Dosher et al., 2020;

Nahum et al., 2010; Yu et al., 2004; Zhang et al., 2008). Moreover, this was true even under some conditions when the references were randomly interleaved across the trials. For example, after a longer, 10-day practice in a bisection discrimination task, participants showed improvement in performance suggesting that PL took place under roving conditions albeit with a reduced magnitude (Parkosadze et al., 2008). This suggested that the lack of learning with multiple references in earlier studies might have reflected a diminished improvement that failed to reach significance rather than a complete elimination of PL. Additional studies revealed that the more perceptually separated the references were from each other in a roving paradigm, the more learning took place. For example, larger frequency difference between the reference sounds in tone discrimination (Amitay et al., 2005), larger angular difference between orientations references in orientation (Dosher et al., 2020; Zhang et al., 2008), and larger contrast difference between the contrast references in contrast discrimination tasks (Zhang et al., 2008), all resulted in more learning than the same training with smaller differences between the references. This indicates an interference effect between the references during learning the discrimination task, which would decrease when the references are well separated along the task-relevant perceptual feature dimension.

In the seminal study of Kuai et al. (2005), the authors found that while learning is disrupted in contrast and motion direction discrimination tasks with randomly interleaved references across trials, if the same references followed a fixed temporal order across trials, PL reemerged (Fig. 4.1B3). Using two different perceptual attributes, these results demonstrated that a temporal regularity of the reference stimuli, other than blocking, can also directly influence PL. Another series of studies showed that learning could take place even with randomly interleaved references if the references in the trials were tagged with symbols indicating the quantity of the reference value relative to the other reference values (Cong & Zhang, 2014; Zhang et al.,

2008). These studies also found that jittering the inter-stimulus-interval (i.e., the time interval between the first and the second stimuli in the temporal 2-AFC task) influenced PL; the larger the randomness in the jitter was, the smaller the improvement on the task became (Zhang et al., 2008).

The studies above suggest that learning is affected by an interference between the references under roving in a complex way. However, these interference effects were reduced or eliminated when the observers were able to distinguish the references from each other. This distinguishing of the references could be achieved in multiple ways: by increasing the difference between the perceptual feature values of the references (Amitay et al., 2005; Dosher et al., 2020; Zhang et al., 2008), by grouping the same references together in blocks (Dosher et al., 2020; Nahum et al., 2010; Tartaglia et al., 2009b; Zhang et al., 2008), by tagging the references with symbols (Cong & Zhang, 2014; Zhang et al., 2008), or by presenting the references in a fixed temporal order (Kuai et al., 2005; Zhang et al., 2008). Supporting the main argument of this thesis, these experiments provide substantial evidence for the interaction of PL and SL by demonstrating that statistical patterns between the references can enable PL. This enabling effect probably emerges due to the structure of the references that enhances the observers' ability to set up contexts and thereby separating the references from each other across the trials.

4.2.2 Existing models and explanations for roving effects

Although there exists a general consensus in the field that the lack of learning in roving conditions can be attributed to interference effects between the references during training, there are several competing models and explanations of the processes that could cause this interference as well as about the factors that modulate the strength of the interference. In this section, I briefly summarize these explanations and emphasize that, beyond the general notion assuming

some sort of top-down basis of the interference during PL, there exists no computational model that can account for the emergence of learning in the "fixed temporal order" condition in roving paradigms.

I start with the reweighting model (Dosher & Lu, 2017), the most successful computational modeling framework in PL to date. This model assumes that there are sensory representational units encoding the stimulus that are pooled together with different weights to decode the correct response corresponding to the stimulus (Fig. 4.3A). In this simple feed-forward architecture, PL emerges by adjusting the decoding weights to increase the accuracy of the response in the task. This model can explain the classical behavioral results showing specificity in PL (Lu et al., 2010; Petrov et al., 2005), however to incorporate the larger set of results showing generalization in some and lack of generalization in some other conditions, the reweighting model needs an extension implemented in the integrated reweighting theory (IRT). In IRT, there are retinal location specific and invariant sensory representational units encoding the feature of stimulus with a location specific and with a location invariant, more abstract representations (Dosher et al., 2013). During the decision process, all of these units are pooled with different weights together to decode the correct response of the trial. In the IRT framework, the interference effects between the reference values in roving is explained by the interfering decoding weights of the location and reference invariant sensory representation layer during learning. The optimization of the decoding weights in the reference invariant layer will suffer from interference effects due to the different optimal decoding weighting for the different references and this interference can explain the lack of learning in roving. In contrast, the emergence of learning in roving can be explained by the co-existence of the location and reference specific representational layers. Since each reference specific layer has its own decoding weights, there will be no interference effects when the weights are optimized to discriminate from multiple reference values and

learning appears in the discrimination task. In sum, the extent of learning in roving depends on the tuning of the decoding weights that pool across the location/reference specific and the location/reference invariant representational layers of the IRT model. However, the influence of the temporal structure in the references (Kuai et al., 2005; Zhang et al., 2008) or the symbolic tagging of the references (Cong & Zhang, 2014; Zhang et al., 2008) cannot be explained in this feed-forward architecture.

The other models explaining the results in roving do not presume specific representations or specific learning processes: these explanations only assume top-down effects modulating the interference between the references in roving tasks. One of the most cited models in PL, the Reverse Hierarchy Theory (RHT) (Ahissar & Hochstein, 2004) posits that learning is gradual in the top-down direction: when the abstract, higher-level structures or the contexts separating the references from each other are learned in roving, only then PL in the lower-level sensory areas will emerge. When there is no learnable structure or context separating the references, PL is disrupted. Another model called the stimulus-tagging model is conceptually very similar to RHT, but it puts the emphasis on the ability to group the references into abstract categories during the training under roving. More specifically, it suggests that in order to avoid the interference effects between the references, the brain needs to tag the stimuli conceptually or semantically into distinct categories so that some top-down attentional processes could switch to the appropriate perceptual template for each reference (Zhang et al., 2008). Finally, yet another model emphasizing top–down connections is based on a reweighting model, in which PL could take place in roving due to the top-down feedback weights being modulated by the learnt task, stimuli structure, and the context (Tartaglia et al., 2009c). Note that our proposed BSPL model provides a conceptually similar explanation for the roving results to the descriptions offered by the models in this paragraph: the statistical structure of the stimuli is learned and it influences the amount

of learning in roving through top-down modulations. However, in contrast to these models, the BSPL model specifies the latent representations involved in learning, it can derive testable predictions and it provides a general learning framework under which PL and SL processes can be treated jointly.

4.3 The Bayesian statistical perceptual learning (BSPL) model

In the previous sections, I pointed out the key finding in the roving paradigm confirmed across several studies: consistent temporal structure between the references enables PL. Existing computational models in the PL literature either cannot explain the emergence of this learning, when temporal pattern is introduced between the references (Dosher et al., 2020), or instead of specifying a concrete learning mechanisms and representation, they propose only unspecified conceptual components, such as a top-down influence (Tartaglia et al., 2009a) or stimulus tagging (Zhang et al., 2008). In this section, I provide a unifying computational model that can explain the influence of the structure in the stimuli along with the other behavior patterns found in both classical and roving PL paradigms by treating PL and SL under the same Bayesian statistical perceptual learning (BSPL) model. The BSPL model is based on the contextual inference model developed in Heald et al. (2020). The key idea behind this framework is combining PL processes found to be responsible for plasticity in early sensory areas with a hidden Markov SL process.

In PL studies the order of the reference- and the test-stimuli is random (see Section 1.1 for a detailed description of a classical PL paradigm), however in order to simplify the generative model of the paradigm that can capture both the classical and the roving studies in PL, I will assume that the reference-stimuli is always presented first and the test-stimuli second after the reference (Fig. 4.2A). Since introducing randomness in the presentation order would only in-

crease uncertainty in PL and SL, but it would not make any systematic deviation in any of the two forms of learning, this simplification of the task does not affect the validity of the BSPL model that I propose.

The rest of this section is structured as follows. First, I will provide the generative model of the BSPL model. Second, using the generative model, I describe how to make inferences about the perceptual feature values of the stimuli assuming temporal dependencies between the reference-stimuli. Finally, I explore a PL mechanism, using a previously developed framework (Ganguli & Simoncelli, 2014), that aims at optimally allocating neurons and spikes given the probability of the feature values of the stimuli by tuning curve modulations, a method of coding frequently reported in neurophysiological studies (LeMessurier & Feldman, 2018; Seriès et al., 2009).

4.3.1 Generative model

Let's consider the following simplified generative model capturing most PL tasks including roving (Fig. 4.2). In each trial, t, the observer's task is to decide whether the value of the reference or the test-stimuli was larger. Therefore, the trials can be sorted into one of the following two decision categories denoted by the binary variable $D_t = \{1, 0\}$: $D_t = 1$ if the value of the reference-stimulus is larger than the value of test-stimulus, and $D_t = 0$ if the reference-stimulus is smaller than the test-stimulus. The decision category of the trials are randomly generated:

$$D_t \sim \text{Bernoulli}\{0.5\} \tag{4.1}$$

In roving, the context/condition of the trial changes across the trials either randomly or following a temporal structure. Here, I address the most widely used roving paradigm, in which there are multiple different reference values and participants are trained to discriminate the

values of the test-stimuli from all the references.

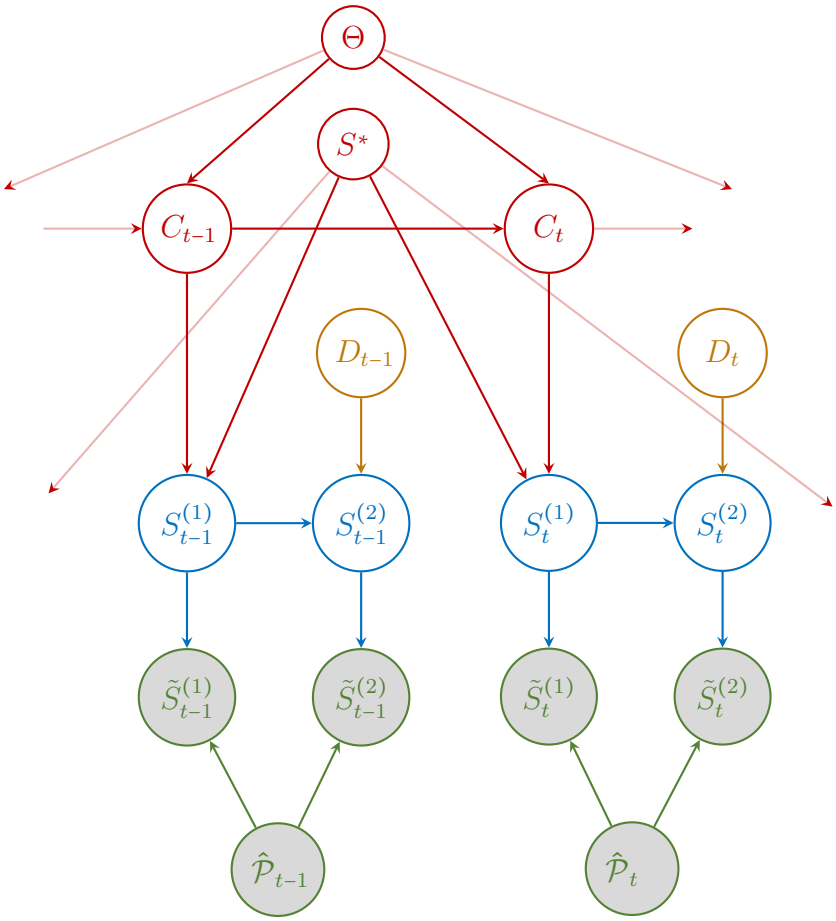


Figure 4.2: The generative scheme of the ideal observer in the Bayesian statistical perceptual learning model. The letters represent random variables and their subscripts, t, denote the trial number. The arrows show the causal dependencies between the variables. In roving paradigms the observer is assumed to learn the context/condition of the trial denoted by a discrete latent variable, C_t. Similar to most roving studies, here, the context of the trials are defined by the reference value in the trial, thus C_t represent the reference-context. The reference-context transitions across the trials following a Markov process governed by the transition model between the reference-contexts represented by Θ. The decision in the trial, representing whether the feature value of the test-stimulus, $S^{(2)}_t$, was larger than the value of the reference-stimuli, $S^{(1)}_t$, is denoted by a binary variable, D_t, colored in gold. The second, test-stimulus is computed by adding an increment to (if $D_t = 1$) or subtracting from (if $D_t = 0$) the reference. The variables representing the stimuli are colored in blue. $\tilde{S}^{(1)}_t$ and $\tilde{S}^{(2)}_t$ denote abstract observations corresponding to the reference- and test- stimuli, respectively. During the SL process the observer learns, Θ, the transition model between the reference-contexts. The variables involved in the SL process are colored in red. During the PL the observer learns to allocate resources in the brain to optimize the sensory encoding of the stimuli for fine discrimination. The variables involved in the PL process are colored in green. The variable, $\hat{\mathcal{P}}_t$, represents the observer's belief about the mean of the distributions over the stimulus encoding models with certain resource allocations. The observer adapts her belief, $\hat{\mathcal{P}}_t$, to optimize performance in the PL task. The observer variables from the observer's perspective are $\tilde{S}^{(1)}_t$, $\tilde{S}^{(2)}_t$, and $\hat{\mathcal{P}}_t$, marked by gray background.

In this paradigm, the context/condition of the trial is defined by the value of the reference and will be called reference-context denoted by C_t. To capture temporal dependency across the reference-contexts in the roving task, I assume that C_t is generated according to a Markov process across the trials, $t = 1, \ldots, T$:

$$C_t \mid C_{t-1}, \{\Theta_i\}_{i=1}^{M} \sim \Theta_{C_{t-1}} \tag{4.2}$$

where $\{\Theta_i\}_{i=1}^{M}$ is the transition probability matrix of the reference-context and Θ_i denotes the ith row capturing the probabilities of transitioning from reference-context i to all possible reference-contexts. The number of possible reference-contexts are assumed to be known by the observer and is denoted by M. The transition probability matrix, $\{\Theta_i\}_{i=1}^{M}$, is learnt by the observer using her observations, and assuming a homogeneous Markov process (see Eqs. 4.15 and 4.16).

In each reference-context i there is a corresponding reference value $S^*_{C_t}$, and the reference-stimulus is generated in a deterministic way given the reference-context and the corresponding reference value:

$$S_t^{(1)} \mid C_t, \{S^*_k\}_{k=1}^{M} \sim \delta\left(S_t^{(1)} - S^*_{C_t}\right) \tag{4.3}$$

where $S_t^{(1)}$ denotes the reference-stimulus, $\{S^*_k\}_{k=1}^{M}$ represents the reference values corresponding to the reference-contexts, $S^*_{C_t}$ is the reference value in reference-context, C_t, and δ denotes a Dirac-delta function. There is a uniform prior distribution over the reference values corresponding to the reference-contexts:

$$S^*_j \sim \mathcal{U}(a, b) \tag{4.4}$$

where a and b are the minimum and maximum values in the perceptual feature space, respectively.

The test-stimulus, $S_t^{(2)}$, is computed by adding an increment to (if $D_t = 0$) or subtracting an increment from (if $D_t = 1$) the reference value. Following the method of constant stimuli in psychophysics (Watson & Fitzhugh, 1990), in each trial, the increment is generated randomly from a uniformly distributed set of increment values:

$$S_t^{(2)} \mid D_t, S_t^{(1)} \sim \begin{cases} \mathcal{U}\left(B,\, S_t^{(1)} + A\right), \text{ if } D_t = 1 \\ \mathcal{U}\left(S_t^{(1)} - A,\, B\right), \text{ if } D_t = 0 \end{cases} \tag{4.5}$$

where $\mathcal{U}$ denotes the uniform distribution and A and B represent the interval, and the smallest value of the increment values, respectively.

Note that several studies in psychophysics implement adaptive methods for measuring discrimination thresholds (García-Pérez, 1998), in which the increment values depend on the previous responses of the observer. Since a possible dependency between previous responses and the current increment would be independent of how the structure in the references influences the PL process (which is the aim of this chapter), it would only make the model more complicated. Therefore, I will use the method of constant stimuli and assume that the increments are generated randomly and they do not depend on the previous responses of the observer.

Finally, $\tilde{S}_t^{(1)}$ and $\tilde{S}_t^{(2)}$ represent the sensory observations corresponding to the reference- and test-stimuli, respectively, formalized as Gaussian random variables:

$$\tilde{S}_t^{(i)} \mid S_t^{(i)}, \hat{\mathcal{P}}_{t-1}(\cdot) \sim \mathcal{N}\left(S_t^{(i)}, \sigma_t^2\{S; \hat{\mathcal{P}}_{t-1}(\cdot)\}\right) \tag{4.6}$$

where $\mathcal{N}$ denotes the Gaussian probability density function with the mean as the first and the variance as the second parameters. During PL the sensory encoding of the perceptual attribute improves. These abstract observation variables in Eq. 4.6 can be related to the neural encoding of the stimulus through the Fisher information which quantifies the amount of information in

a neuronal population activity in response to a stimulus value (see Appendix C and Dayan and Abbott, 2005 for more information on the Fisher information). If one assumes that the population response can be captured by independent Poisson random variables, then, in the limit of large number of neurons, the variance of the maximum likelihood estimate of the stimulus value is inversely proportional to the Fisher information (Dayan & Abbott, 2005). Intuitively, the more information there is in the population response about the stimulus value the smaller the uncertainty becomes when estimating the stimulus value. Therefore, the variance in 4.6 can be written using the amount of Fisher information in the response of a hypothetical population of sensory neurons:

$$\sigma_t^2\{S;\hat{\mathcal{P}}_{t-1}(\cdot)\} = \frac{1}{I_{\mathrm{F}}\{S;\hat{\mathcal{P}}_{t-1}(\cdot)\}} \tag{4.7}$$

I_{F} is the Fisher information which is a function of the stimulus value, S, with perceptual encoding parameters, optimized using all the observations until the previous trial denoted by $\hat{\mathcal{P}}_{t-1}(\cdot)$ (see more information about $\hat{\mathcal{P}}_{t-1}(\cdot)$ in Section 4.3.4). Using the formulation above, the variance in Eq. 4.6 also becomes a function of the stimulus values with the same perceptual parameters which create the dependence on the previous observations, $\tilde{S}^{(1)}_{1:t-1}, \tilde{S}^{(2)}_{1:t-1}$, on the left side of 4.6 through $\hat{\mathcal{P}}_{t-1}(\cdot)$. Section 4.3.3 will provide a detailed description about the perceptual encoding parameters and how the observer optimizes encoding to increase performance in the task.

4.3.2 Inference

Since the participants cannot observe directly the values of the stimuli nor how the values of the reference change in the trials, they have to infer it from their sensory observations. In this section, I will describe how the values of the reference, the test, and the decision category of the trials are inferred based on the sensory observations.

Inference over the reference-context: Due to the temporal structure in the reference-contexts the inference is based on a first order hidden Markov model (see Fig. 4.2) in which the observer infers the value of the reference-context at the current trial based on all her sensory observations until that trial (Rabiner, 1989).

First, the joint probabilities of the previous and the current reference-contexts will be formalized which will be used during the SL process in Eqs. 4.23:

$$\begin{aligned} &\mathcal{P}\left(C_{t-1}=i, C_t=j \mid \tilde{S}^{(1)}_{1:t}, \tilde{S}^{(2)}_{1:t}\right) \simeq \pi_{ij}(t) \\ &\pi_{ij}(t) = \frac{\mathcal{P}\left(\tilde{S}^{(1)}_t, \tilde{S}^{(2)}_t \mid C_t=j, \tilde{S}^{(1)}_{1:t-1}, \tilde{S}^{(2)}_{1:t-1}\right) \hat{\Theta}_{ij}(t) \sum_{k=1}^{M} \pi_{ki}(t-1)}{\sum_{l,m=1}^{M} \mathcal{P}\left(\tilde{S}^{(1)}_t, \tilde{S}^{(2)}_t \mid C_t=m, \tilde{S}^{(1)}_{1:t-1}, \tilde{S}^{(2)}_{1:t-1}\right) \hat{\Theta}_{lm}(t) \sum_{k=1}^{M} \pi_{kl}(t-1)} \end{aligned} \tag{4.8}$$

where $\pi_{ij}(t)$ represents the current approximated estimate of the joint probability of i being the previous and j being the current reference-contexts, $\hat{\Theta}_{i,j}(t)$ denotes the current approximated estimate of the transition probability from the ith to the jth context, and the subscripts, $1:t$ and $1:t-1$, denote all the trials including or excluding the current trial t, respectively.

Second, the probability of the reference-context at the current trial can be computed by marginalizing the joint probability in 4.8 over the previous reference-context:

$$\mathcal{P}\left(C_t=j \mid \tilde{S}^{(1)}_{1:t}, \tilde{S}^{(2)}_{1:t}\right) \simeq \sum_{k=1}^{M} \pi_{kj}(t) \tag{4.9}$$

The probability of the reference-context at the current trial in Eq. 4.9 can be interpreted as the responsibility that a reference-context j takes for explaining the current observations (see the term responsibility in mixture models, e.g., in Bishop, 2006).

Finally, the predictive probability of the references, which will be important in the PL algorithm (see the algorithm in Section 4.3.4), can be computed from the transition probability matrix and responsibility from the previous trial (i.e., the marginal probability of the previous

reference-context):

$$\mathcal{P}\left(C_t = j \mid \tilde{S}^{(1)}_{1:t-1}, \tilde{S}^{(2)}_{1:t-1}\right) \simeq \sum_{l=1}^{M} \hat{\Theta}_{lj}(t-1)\, \mathcal{P}\left(C_{t-1} = l \mid \tilde{S}^{(1)}_{1:t-1}, \tilde{S}^{(2)}_{1:t-1}\right) \tag{4.10}$$

The likelihood term on the right side of Eq. 4.8 still needs to be specified:

$$\begin{aligned}
&\mathcal{P}\left(\tilde{S}^{(1)}_t, \tilde{S}^{(2)}_t \mid C_t = j, \tilde{S}^{(1)}_{1:t-1}, \tilde{S}^{(2)}_{1:t-1}\right) = \\
&\iint \mathcal{P}\left(\tilde{S}^{(1)}_t \mid S^{(1)}_t, \hat{\mathcal{P}}_{t-1}(\cdot)\right) \mathcal{P}\left(\tilde{S}^{(2)}_t \mid S^{(2)}_t, \hat{\mathcal{P}}_{t-1}(\cdot)\right) \\
&\sum_{d=0}^{1} \mathcal{P}\left(S^{(2)}_t \mid S^{(1)}_t, D_t\right) \mathcal{P}(D_t = d) \\
&\int \mathcal{P}\left(S^{(1)}_t \mid C_t = j, \{S^*_k\}\right) \mathcal{P}\left(\{S^*_k\} \mid \tilde{S}^{(1)}_{1:t-1}, \tilde{S}^{(2)}_{1:t-1}\right) \mathrm{d}\{S^*_k\}\, \mathrm{d}S^{(1)}_t\, \mathrm{d}S^{(2)}_t
\end{aligned} \tag{4.11}$$

Using the generative model described in Section 4.3.1 and in Fig. 4.2 the equation above can be simplified and written in terms of the probability distributions corresponding to the the variables:

$$\begin{aligned}
&\mathcal{P}\left(\tilde{S}^{(1)}_t, \tilde{S}^{(2)}_t \mid C_t = j, \tilde{S}^{(1)}_{1:t-1}, \tilde{S}^{(2)}_{1:t-1}\right) = \\
&\iint \mathcal{N}\left\{\tilde{S}^{(1)}_t; S^{(1)}_t, \sigma^2\left(S^{(1)}_t; \hat{\mathcal{P}}_{t-1}(\cdot)\right)\right\}\, \mathcal{N}\left\{\tilde{S}^{(2)}_t; S^{(2)}_t, \sigma^2\left(S^{(2)}_t; \hat{\mathcal{P}}_{t-1}(\cdot)\right)\right\} \\
&\sum_{d=0}^{1} \mathcal{U}\left\{S^{(2)}_t; D, E\right\} \mathcal{P}(D_t = d) \\
&\mathcal{N}\left\{S^{(1)}_t; \mu^{(j)}_{t-1}, \omega^{(j)}_{t-1}\right\}\, \mathrm{d}S^{(1)}_t \mathrm{d}S^{(2)}_t
\end{aligned} \tag{4.12}$$

where $\mu^{(j)}_{t-1}$ and $\omega^{(j)}_{t-1}$ denote the inferred mean and the variance of the probability distribution over the reference values in the j context from the previous trial (see their derivations in the next section). D and E denote the range of the uniform distribution depending on the decision category of the trials (see Eq. 4.5). Calculating Eq. 4.12 is still intractable and I will approximate it by evaluating the integrals on a discrete grid in the stimuli feature space.

Inference over the reference value in the contexts: The observer also infers the value of the reference in the contexts using her observations. Since the prior over the reference values is uniformly distributed (Eq. 4.4), the sensory observations given the reference-stimuli has a Gaussian distribution (Eq. 4.6), and the reference-stimulus given the reference value is a Dirac delta (Eq. 4.3) the posterior distributions over the reference values associated with the contexts will be Gaussian distributions:

$$\mathcal{P}\left(S_j^* \mid \tilde{S}_{1:t}^{(1)}, \tilde{S}_{1:t}^{(2)}\right) = \mathcal{N}\left(S_j^*; \mu_t^{(j)}, \omega_t^{(j)}\right) \tag{4.13}$$

One can write an update rule for the mean and the variance in the following way:

$$\bar{\omega}_t^{(j)} = \bar{\omega}_{t-1}^{(j)} + \frac{\mathcal{P}\left(C_t = j \mid \tilde{S}_{1:t}^{(1)}, \tilde{S}_{1:t}^{(2)}\right)}{\sigma_t^2\left(S_t^{(1)}; \hat{\mathcal{P}}_t(\cdot)\right)} \tag{4.14}$$

$$\bar{\mu}_t^{(j)} = \bar{\mu}_{t-1}^{(j)} + \frac{\mathcal{P}\left(C_t = j \mid \tilde{S}_{1:t}^{(1)}, \tilde{S}_{1:t}^{(2)}\right)}{\sigma_t^2\left(S_t^{(1)}; \hat{\mathcal{P}}_t(\cdot)\right)} \tilde{S}_t^{(1)} \tag{4.15}$$

$$\omega_t^{(j)} = \frac{1}{\bar{\omega}_t^{(j)}} \tag{4.16}$$

$$\mu_t^{(j)} = \omega_t^{(j)} \bar{\mu}_t^{(j)} \tag{4.17}$$

Using this update, only the previous values of the mean and the variance are stored and get updated by the responsibility and the observations in the current trial.

Inference over the reference- and test-stimulus: The observer needs to infer the reference- and the test stimuli to make a decision about the trial's category and to reallocate the neural resources to improve the encoding of the stimuli for the discrimination during PL. From the joint probability of the reference- and test-stimuli one can compute both the decision category

of the trial (see Eqs. 4.20 and 4.21) and the optimal resource allocation (see Eq. 4.27):

$$\begin{aligned}
&\mathcal{P}\left(S_t^{(1)}, S_t^{(2)} \mid \tilde{S}_{1:t}^{(1)}, \tilde{S}_{1:t}^{(2)}\right) \propto \\
&\mathcal{P}\left(\tilde{S}_t^{(1)} \mid S_t^{(1)}, \hat{\mathcal{P}}_{t-1}(\cdot)\right) \mathcal{P}\left(\tilde{S}_t^{(2)} \mid S_t^{(2)}, \hat{\mathcal{P}}_{t-1}(\cdot)\right) \\
&\sum_{d=0}^{1} \mathcal{P}\left(S_t^{(2)} \mid S_t^{(1)}, D_t\right) \mathcal{P}(D_t = d) \\
&\sum_{j=1}^{M} \mathcal{P}\left(C_t = j \mid \tilde{S}_{1:t-1}^{(1)}, \tilde{S}_{1:t-1}^{(2)}\right) \int \mathcal{P}\left(S_t^{(1)} \mid C_t = j, \{S_k^*\}\right) \mathcal{P}\left(\{S_k^*\} \mid \tilde{S}_{1:t-1}^{(1)}, \tilde{S}_{1:t-1}^{(2)}\right) \mathrm{d}\{S_k^*\}
\end{aligned} \tag{4.18}$$

Since the computing the joint probability above is intractable approximations are needed to evaluate it. During the approximation the term that normalizes the joint probability can be ignored (note the proportional sign in Eq. 4.18) and the normalization will be approximated numerically. In the same way as in Eq. 4.12 the equation above can be simplified and written in terms of the probability distributions of the variables:

$$\begin{aligned}
&\mathcal{P}\left(S_t^{(1)}, S_t^{(2)} \mid \tilde{S}_{1:t}^{(1)}, \tilde{S}_{1:t}^{(2)}\right) \propto \\
&\mathcal{N}\left\{\tilde{S}_t^{(1)}; S_t^{(1)}, \sigma^2\left(S_t^{(1)}; \hat{\mathcal{P}}_{t-1}(\cdot)\right)\right\} \mathcal{N}\left\{\tilde{S}_t^{(2)}; S_t^{(2)}, \sigma^2\left(S_t^{(2)}; \hat{\mathcal{P}}_{t-1}(\cdot)\right)\right\} \\
&\sum_{d=0}^{1} \mathcal{U}\left\{S_t^{(2)}; D, E\right\} \mathcal{P}(D_t = d) \\
&\sum_{j=1}^{M} \mathcal{P}\left(C_t = j \mid \tilde{S}_{1:t-1}^{(1)}, \tilde{S}_{1:t-1}^{(2)}\right) \mathcal{N}\left\{S_t^{(1)}; \mu_{t-1}^{(j)}, \omega_{t-1}^{(j)}\right\}
\end{aligned} \tag{4.19}$$

Inference over the trial's decision category: The posterior probability of the trial's category can be computed directly from the joint probability of the reference- and test-stimuli in Eq. 4.19 by computing the probability of the difference between the reference- and test-stimuli:

$$\begin{aligned}
\mathcal{P}(D_t = 1 \mid \tilde{S}_{1:t}^{(1)}, \tilde{S}_{1:t}^{(2)}) = \mathcal{P}(S_t^{(1)} - S_t^{(2)} > 0 \mid \tilde{S}_{1:t}^{(1)}, \tilde{S}_{1:t}^{(2)}) \\
\int_0^{\mathrm{Max}} \mathcal{P}(S_t^{(1)} - S_t^{(2)} \mid \tilde{S}_{1:t}^{(1)}, \tilde{S}_{1:t}^{(2)})
\end{aligned} \tag{4.20}$$

Let $Z_t = S_t^{(1)} - S_t^{(2)}$, denoting the difference of the two stimuli values, then the probability distribution over the difference can be given by convolving the joint posterior probability distributions of the two stimuli:

$$\mathcal{P}(Z_t = z \mid \tilde{S}_{1:t}^{(1)}, \tilde{S}_{1:t}^{(2)}) = \int \mathcal{P}\left(S_t^{(1)} = s, S_t^{(2)} = s - z \mid \tilde{S}_{1:t}^{(1)}, \tilde{S}_{1:t}^{(2)}\right) \mathrm{d}s \tag{4.21}$$

4.3.3 Statistical learning

After describing the inference in the BSPL model, I turn to capturing the learning processes during the task. In the roving paradigm the observer is assumed to track the statistics of the reference-contexts by updating her belief about the transition probabilities between the reference-contexts after each trial. The statistically optimal learning model, derived in Rabiner (1989), shows that the optimal transition probability matrix, Θ, can be computed as the expected number of transitions from reference-context i to j across the trials divided by the expected number of transitions from reference-context i. I will use the following iterative and recursive approximation of the statistically optimal learning:

$$\hat{\Theta}_{ij}(t) = \frac{\hat{\pi}_{ij}(t) + \sum_{\tau=1}^{t-1} \hat{\pi}_{ij}(\tau)}{\sum_{k=1}^{M} \hat{\pi}_{ik}(t) + \sum_{\tau=1}^{t-1} \hat{\pi}_{ik}(\tau)} \tag{4.22}$$

$$\hat{\pi}_{i,j}(t) = \frac{\mathcal{P}\left(\tilde{S}_t^{(1)}, \tilde{S}_t^{(2)} \mid C_t = j, \tilde{S}_{1:t-1}^{(1)}, \tilde{S}_{1:t-1}^{(2)}\right) \hat{\Theta}_{ij}(t) \sum_{k=1}^{M} \hat{\pi}_{ki}(t-1)}{\sum_{lm=1}^{M} \mathcal{P}\left(\tilde{S}_t^{(1)}, \tilde{S}_t^{(2)} \mid C_t = m, \tilde{S}_{1:t-1}^{(1)}, \tilde{S}_{1:t-1}^{(2)}\right) \hat{\Theta}_{lm}(t) \sum_{k=1}^{M} \hat{\pi}_{kl}(t-1)} \tag{4.23}$$

In order to approximate the optimal transition probability matrix $\hat{\Theta}_{ij}(t)$ and $\hat{\pi}_{ij}(t)$ should be updated iteratively until convergence. For the first trial, I assume that the observer uses equal initial probabilities of observing any of the references-contexts. After the first trial, in each trial, at the first iteration, the transition probability matrix from the previous trial is used. This iterative approximation procedure can be interpreted as an approximation of an Expectation

Maximization algorithm (Dempster et al., 1977) for hidden Markov models (Baum, 1972).

4.3.4 Perceptual learning

In the previous sections I described the inference and the SL process in the BSPL model. In this section I will explain how perceptual learning can be added to this unifying framework.

The PL mechanisms will be introduced into the BSPL model using the encoding-decoding framework (Ganguli & Simoncelli, 2014; Seriès et al., 2009; Stocker & Simoncelli, 2006; Wei & Stocker, 2015). Perception has two stages in the encoding-decoding scheme. First, at the encoding stage, the sensory representation of the stimuli is created based on the stimulus-dependent activity of the sensory units. This stage is formalized by the likelihood of the stimulus's feature value $\mathcal{P}(r_t^{(i)} \mid S_t^{(i)})$ where $r_t^{(i)}$ represents the neural response to the two stimuli and $S_t^{(i)}$ denotes the feature values of those stimuli. Second, at the decoding stage, a belief about the true value of the stimulus's feature is estimated from the sensory representation. An optimal decoder computes the posterior probability of the feature value based on the true generative model of the sensory representation using Bayes' rule: $\mathcal{P}(S_t^{(i)} \mid r_{1:t}^{(i)}) \propto \mathcal{P}(r_t^{(i)} \mid S_t^{(i)})\mathcal{P}(S_t^{(i)} \mid r_{1:t-1}^{(i)})$.

PL has been connected to a wide range of neural areas and neural learning effects (see the details and the references in Section 1.1.2), thus several learning mechanisms were proposed to account for the improvement in fine discrimination during PL tasks (see Dosher and Lu, 2017; LeMessurier and Feldman, 2018; Schwabe, 2005; Teich and Qian, 2003). Under the encoding-decoding framework, these mechanisms can broadly be grouped into two categories. Models in the first category assumes that the neural representation (i.e., the encoding) of the trained stimulus in early sensory areas improves due to learning (e.g., LeMessurier and Feldman, 2018; Seriès et al., 2009) while the models in the second category propose that learning only adapts the readout (i.e. the decoding) from the early sensory areas and other higher-level,

decision-making processes (e.g., Dosher et al., 2013). More recently researchers have argued that learning probably takes place at multiple brain areas and involves multiple mechanisms (LeMessurier & Feldman, 2018; Maniglia & Seitz, 2018; Watanabe & Sasaki, 2015).

Since multiple mechanisms are able to capture the same behavior in most PL tasks and the aim of this chapter is to demonstrate how to treat PL and SL in one unifying computational framework rather than comparing the existing PL mechanisms or propose a novel PL mechanism I will only explore one framework here. Furthermore, the data, consisting of average performance levels, from the behavior studies, that I aim to explain jointly with the BSPL model, is not sufficient to distinguish between the PL mechanisms proposed in the literature (Dosher & Lu, 2017; LeMessurier & Feldman, 2018; Schwabe, 2005; Teich & Qian, 2003). For these reasons, I chose to explore a framework (Ganguli & Simoncelli, 2014) that can seamlessly be integrated in the BSPL model and that offers closed-form solutions for maximizing performance in discrimination with both Gaussian-like and sigmoidal tuning curves.

Implementing optimal encoding for multiple references: Ganguli and Simoncelli (2014) derived a closed-form solution for optimally allocating sensory neurons and spikes to maximize discrimination performance given a prior distribution over the stimulus values and some resource constraints. Similar to other studies (Jazayeri & Movshon, 2006; Seriès et al., 2009; Seung & Sompolinsky, 1993), the authors assumed that there is a single, homogeneous population of sensory neurons with unimodal or sigmoidal response profiles tuned to different feature values spanning the whole feature space uniformly. The number of spikes the neurons emitted in a given period of time was generated from independent Poisson distributions with mean activities described by the tuning curves of the neurons. The number of neurons and the total expected spike rate of the neuron population were assumed to be fixed constraining the tuning curves. Given these assumptions the authors asked how to encode the stimulus efficiently

by changing the density and the gain of the neurons along the feature space given a resources budget (i.e., the number of neurons and spikes). They derived that the optimal solution for allocating the resources to maximize discrimination performance only depends on the probability distribution over the stimulus feature values. The solution, intuitively, was to allocate more neurons for feature values that are more probable in the task which will result in lower discrimination thresholds (see the detailed derivation in Ganguli and Simoncelli, 2014) and a short description in Appendix C.3).

Furthermore, the authors gave a closed-form solution for computing the Fisher information in the response of a population whose resources were optimized to maximize discrimination performance. It turned out that the Fisher information also depends solely on the probability distribution of the stimulus feature values. The BSPL model is connected to neural coding using the Fisher information in Eq. 4.7. Therefore, a PL process that optimize the encoding of the stimulus feature values by adapting the density and the gain of the neurons along the feature space given a resources budget can be implemented by (1) updating the observer's belief about the probability distribution over the stimulus feature values presented in the experiment and (2) computing the Fisher information given that probability distribution over the stimulus values.

The closed-form solution for computing the Fisher information for the population optimized for encoding the stimulus values in the experiment given their probability distribution is the following:

$$I_{\mathrm{F}}\left\{S, \hat{\mathcal{P}}(\cdot)\right\}_{\text{Gaussian}} \propto N^2 \sqrt{\hat{\mathcal{P}}(S)} \tag{4.24}$$

$$I_{\mathrm{F}}\left\{S, \hat{\mathcal{P}}(\cdot)\right\}_{\text{Sigmoid}} \propto N \hat{\mathcal{P}}^{\frac{2}{3}}(S) \left[1 - \int_{-\infty}^{S} \hat{\mathcal{P}}(S)\, \mathrm{d}S\right]^{-\frac{1}{3}} \tag{4.25}$$

where S denotes the stimulus feature values, $\hat{\mathcal{P}}(\cdot)$ represent the probability distribution of the stimulus feature values, and N is the resource constraint (the number of neurons and spikes

together). Eq. 4.24 is the solution for unimodal, Gaussian-like while Eq. 4.25 is the solution for sigmoidal tuning curves. These equations (4.24 and 4.25) show how to compute the Fisher information for a population response whose resources were already optimized for discrimination in terms of the location of the tuning preferences and number of spikes given the probability of the stimulus feature values. Therefore, to implement learning in the BSPL model one only needs to update the probability distribution over the stimulus feature values, $\hat{\mathcal{P}}(S)$, during the experiment.

Moreover, the Fisher information can be used to derive the minimum achievable discrimination thresholds during the PL task (Seriès et al., 2009). Since the Fisher information for a population response whose resources were already optimized for discrimination can be computed from the probability distribution over the stimulus feature values, $\hat{\mathcal{P}}(S)$, the lower bound of the discrimination threshold can also be given using $\hat{\mathcal{P}}(S)$ only:

$$\Delta\left\{S, \hat{\mathcal{P}}(\cdot)\right\}_{\text{Gaussian}} \propto \hat{\mathcal{P}}^{-\frac{1}{4}}(S) \tag{4.26}$$

$$\Delta\left\{S, \hat{\mathcal{P}}(\cdot)\right\}_{\text{Sigmoid}} \propto \hat{\mathcal{P}}^{-\frac{1}{3}}(S)\left[1 - \int_{-\infty}^{S} \hat{\mathcal{P}}(S)\,\mathrm{d}S\right]^{\frac{1}{6}} \tag{4.27}$$

where $\Delta\left\{S, \hat{\mathcal{P}}(\cdot)\right\}$ denotes the lower bound on the discrimination threshold.

However, the BSPL model allows to have multiple references. Therefore, the objective of PL is to adjust the encoding of the stimuli through optimal resource allocation to improve the discrimination performance from multiple references. Optimizing encoding for multiple references will result in interference effects. The optimal resource allocation for one reference puts most resources to improve the encoding of feature values that are not important for another reference and an optimal encoding would allocate minimal, or no resources for those feature

values in case of that other reference. Thus, the adjusting for both references could cancel each other out or reduce the potential learning. In most roving conditions, there are multiple (i.e., more than two) references evenly distributed across the range of possible feature values, which leads to interference effects that decreases the performance evenly for all of the references. Indeed, most computational models and explanations for roving in the literature explain the lack of learning in roving conditions with an interference effect between the references (see section 4.2.2 and Dosher et al., 2020; Nahum et al., 2010; Tartaglia et al., 2009c; Zhang et al., 2008).

However, there are conditions in roving (e.g., blocked and fixed order, see section 4.2.1 and Fig. 4.1B) in which PL emerges and conditions (e.g., randomly interleaved, see section 4.2.1 and Fig. 4.1B) in which PL is disrupted. How can learning in one condition and the lack of learning in other conditions be explained? Most accounts in the reweighting framework assume that there are both condition specific decoding weights, contributing to learning in all conditions without interference, and condition invariant decoding weights, resulting in an interference effect between the conditions (Dosher & Lu, 2017; Dosher et al., 2020; Talluri et al., 2015). Therefore, the emergence of learning in fine discrimination with multiple references can be explained with reference specific parameters while the interference effects between the references can originate from reference invariant/general parameters.

In the BSPL model, to improve discrimination performance from all references without or with reduced interference the observer should use and adapt different encoding models and efficiently allocate the resources for discrimination with multiple references. This process can be framed as causal learning (Jacobs & Kruschke, 2011; Körding et al., 2007) in which each reference-context corresponds to a different latent cause that could have generated the sensory

observations.

First, during this causal inference in the BSPL model, the observer infers the predictive probability of the different latent causes, i.e., the predictive probability of the reference-contexts given the observations until the previous trials, but before observing the stimulus in the current trial (Eq. 4.10). Then, the observer prepares for perceiving the stimuli in the current trial by allocating the resources for optimal encoding given the predictive probability of the reference-contexts. During this process she combines the encoding models, that allocate their resources efficiently to the different latent causes (i.e., reference-contexts), with weights proportional to the predictive probability of the different latent causes:

$$\hat{\mathcal{P}}_{t-1}(S) = \sum_{j=1}^{M} \hat{\mathcal{P}}_{t-1}^{(j)}(S)\, \mathcal{P}\left(C_t = j \mid \tilde{S}_{1:t-1}^{(1)}, \tilde{S}_{1:t-1}^{(2)}\right) \tag{4.28}$$

where $\hat{\mathcal{P}}_{t-1}^{(j)}(S)$ denotes the predicted probability of the feature values of the stimuli given the encoding model with the optimal resource allocation corresponding to the jth reference-context. Since only the Fisher information is needed to compute sensory likelihoods which information depends only on the probability of the stimulus values this causal inference process can be implemented solely by tracking and updating separate probability distributions over the feature values of stimuli corresponding to the different references-contexts (hidden causes). Therefore, during PL only the probability distributions over the stimulus values associated with the reference-contexts will be updated:

$$\hat{\mathcal{P}}_{t}^{(j)}(S) = (1-\alpha)\, \hat{\mathcal{P}}_{t-1}^{(j)}(S) + \alpha\, \mathcal{P}\left(C_t = j \mid \tilde{S}_{1:t}^{(1)}, \tilde{S}_{1:t}^{(2)}\right) q_t \tag{4.29}$$

$$q_t = (1-\gamma) \int \mathcal{P}\left(S_t^{(1)} = \tilde{s} \mid \tilde{S}_{1:t}^{(1)}, \tilde{S}_{1:t}^{(2)}\right) \kappa(\tilde{s} - S)\, \mathrm{d}\tilde{s} + \gamma \int \mathcal{P}\left(S_t^{(2)} = \tilde{s} \mid \tilde{S}_{1:t}^{(1)}, \tilde{S}_{1:t}^{(2)}\right) \kappa(\tilde{s} - S)\, \mathrm{d}\tilde{s} \tag{4.30}$$

$\hat{\mathcal{P}}(S)$ represent the mean of the probability distribution over the probability distributions of the stimulus feature values associated with the reference-contexts. The mean over the distributions of the stimulus values corresponding to the jth reference-context is updated by combining the probability of the feature values of the stimuli in the previous trial, $\hat{\mathcal{P}}^{(j)}_{t-1}(S)$ with the probability of the stimuli in the current trial given all observations, q_t. α denotes a learning rate that determines the weighting between the probability distribution of the feature values in the previous and in the current trials. If one assumes a static Dirichlet distribution prior over $\hat{\mathcal{P}}(S)$ then the mean of the posterior Dirichlet distribution can be given by Eq. 4.29 with $\alpha = \frac{1}{\beta_0 + t}$ where β_0 is the parameter of the prior Dirichlet distribution over $\hat{\mathcal{P}}(S)$.

In Eq. 4.30 the probability of the stimulus feature values combines the probability of the reference- and test-stimuli in the current trial with a weight denoted by γ. This parameter can determine whether the reference- or the test-stimuli require more resources in the encoding to improve performance in the discrimination. The optimal solution would be to allocate more resources to the stimuli with more uncertainty. Since in most PL paradigms the same reference-stimulus is observed in every trial while the test-stimuli changes across the trials the uncertainty in the reference-stimulus is much lower than in the test-stimulus. Even in the roving paradigms the reference-stimuli repeats itself in every 4 trials while the test-stimuli repeats itself much less frequently. Therefore, an optimal observer would allocate more resources to encode the test-stimuli in PL paradigms which can be formalized with $\gamma > 0.5$.

Finally, if S is different than the feature space of the stimuli the probabilities of the stimuli in the current trial, $\mathcal{P}\left(S^{(1)}_t = \tilde{s} \mid \tilde{S}^{(1)}_{1:t}, \tilde{S}^{(2)}_{1:t}\right)$ and , $\mathcal{P}\left(S^{(2)}_t = \tilde{s} \mid \tilde{S}^{(1)}_{1:t}, \tilde{S}^{(2)}_{1:t}\right)$ can be related to $\hat{\mathcal{P}}^{(j)}_{t-1}(S)$ with a smooth kernel function κ (e.g., with a squared exponential kernel function: $\kappa(r) = e^{-\frac{r}{2l^2}}$ where l is the characteristic length-scale (Rasmussen, 2004)).

The learning algorithm in the BSPL model: In this section, I write down the learning algorithm used in the BSPL model.

Step 0: Bring the responsibility from the previous trial, $\mathcal{P}\left(C_{t-1} = j \mid \tilde{S}^{(1)}_{1:t-1}, \tilde{S}^{(2)}_{1:t-1}\right)$.

Step 1: Compute the predictive probability of the reference-contexts, $\mathcal{P}\left(C_t = j \mid \tilde{S}^{(1)}_{1:t-1}, \tilde{S}^{(2)}_{1:t-1}\right)$, using Eq. 4.10.

Step 2: Optimize the resource allocation in the stimulus encoding using the predictive probability of the reference-contexts and compute $\hat{\mathcal{P}}_{t-1}(S)_{\text{predictive}}$ using Eq. 4.28.

Step 3: Compute the Fisher information, $I_{\text{F}}\left\{\hat{\mathcal{P}}_{t-1}(S)_{\text{predictive}}\right\}$, using Eq. 4.24 or Eq. 4.25.

Step 4: Update the transition probability matrix of the references, $\hat{\Theta}_{ij}(t)$, given the observation model in 4.6 with Eqs. 4.22 and 4.23.

Step 5: Compute the joint probability of the reference- and test-stimuli given the observation model in 4.6 and $\hat{\Theta}_{ij}(t)$ with Eq. 4.19.

Step 6: Compute the decision category of the trial, D_t using Eqs. 4.20 and 4.21 or Eqs. 4.26 and 4.27.

Step 7: Compute the context responsibilities, $\mathcal{P}\left(C_t = j \mid \tilde{S}^{(1)}_{1:t}, \tilde{S}^{(2)}_{1:t}\right)$, based on the observations using Eq. 4.9.

Step 8: Update the probability distributions over the stimulus values associated with the reference-contexts, $\hat{\mathcal{P}}^{(j)}_t(S)$, given the responsibility in the current trial and the probability of the reference- and test-stimuli using Eqs. 4.29 and 4.30

In this algorithm the observer allocates her resources in the encoding of both the reference- and the test-stimuli based on the predictive probability of the reference values (steps 2-6). Another modelling choice could be to assume that the observer updates her resource allocation after

observing the reference-stimuli and compute $\hat{\mathcal{P}}_t^{(j)}(S)$ before step 4 and recalculate the resource allocation and the Fisher information in step 2 and 3 before calculating the joint probability in step 5.

The interaction between PL and SL: SL is the process by which the observer makes inference about the reference-context and -value in the trials, while PL is the process during which the encoding of the stimuli gets adapted to the structure in the stimuli to increase the performance in the task.

SL influences performance in two ways. There is an effect of the prior distribution over the reference-contexts quantified by the term, $\sum_{k=1}^{M} \pi_{ki}(t-1)$, in Eq. 4.8. Inferring the reference-contexts also determines how to allocate the resources for the stimulus encoding during perception (using Eqs. 4.26, 4.24 or 4.25, and 4.6) and learning (Eq. 4.28). The effect of the prior distribution over the references is much smaller than the effect of choosing the right encoding model for the reference-contexts in the trials. Intuitively, even if the prior expectation about the reference-context is wrong the observer will see the reference-stimulus in the trial which provides huge amount of information about the reference value and the context. However, if the observer allocates her resources for the wrong reference-context, then the stimulus encoding could be inappropriate for the true reference- and test-stimuli in the trial resulting in bad sensory observations that not just cannot compensate for the wrong prior expectation, but it could make the inference worse.

PL also influences SL through the adaptation of the encoding models in Eq. 4.25. The more the encoding model adapts to maximize discrimination between the reference-contexts the better the observer can differentiate between the reference-stimuli accelerating SL.

4.4 Simulation results

In this section, I will replicate the pattern of results found in empirical studies using roving (see section 4.2.1) by generating synthetic responses probabilities from the BSPL model (using Eq. 4.20). Simulating behavioral results is an important first step in demonstrating that the BSPL model is the correct framework for capturing the pattern of human behavioral results found in the literature. It will be for future studies to close the loop by conducting experiments with human observers and carrying out a rigorous model fitting to their responses for the ultimate validation of the model. Here, I will account for the average discrimination threshold changes due to learning because that was the measured dependent variable in all of the previously conducted roving experiments.

4.4.1 Details of the simulation

The two most frequently used perceptual attributes in roving studies were contrast and orientation, and the BSPL model can be used for both attributes using the closed-form solutions in Eqs. 4.24 - 4.25. I will demonstrate the model's performance using only orientation discrimination, but it is straightforward to implement the model for the contrast discrimination task too.

The probabilities of the trials' decision category was generated using the BSPL model (see Section 4.3.4, step 6) in four hypothetical orientation discrimination experiments copying the the classical PL paradigm with one reference and the three main conditions in roving, the randomly interleaved, the blocked, and the fixed order conditions (see section 4.2.1 and Fig. 4.1B). Based on Kuai et al. (2005) the following four reference values were used: 22.5°, 67.5°, 112.5° and 157.5°. To assess discrimination thresholds the method of constant stimuli (Watson & Fitzhugh, 1990) were applied with ten increment values: 1°, 1.39°, 1.94°, 2.71°, 3.78°, 5.27°, 7.35°, 10.25°, 14.30°, 19.95°. When the increments are chosen properly, the observer's true

threshold lies at the midpoint of the range of the increments (note that for most perceptual attributes a logarithmic scale is used for the increments). The specific increment values were chosen based on our study (Lengyel & Fiser, 2019) in which most observers had a threshold between 5°-15° in the orientation discrimination tasks.

In each experiment, there was a 5-day training procedure. Each day ten repetitions with each increment values per reference value were presented in a random order (20 test-stimuli [2×10 increments] × 4 references × 10 = 800 trials per day). In the single experiment there was only one, single reference in the all of the trials. In the randomly interleaved experiment, the reference values were randomly generated in each trial from the four values. In the blocked experiment, the 200 trials with same references ([2×10 increments] × 1 reference × 10 = 200) were grouped into separate blocks. In the fixed order experiment, the reference changed in every trial, however their presentation sequence followed a fixed order during the training.

Since the generation of the hypothetical observations in Eq. 4.24 is stochastic I performed the generation of the decision category probabilities ten times in each experiment and took the mean of the 10 decision category probabilities for each increment value. The variance over the decision category probabilities were smaller than 0.003 in each experiment, therefore only the mean values are shown in Figs. 4.3 - 4.6.

4.4.2 Results of the simulation

In all simulated experiments, PL showed its typical markers. First, performance decreased monotonically with smaller increment values resulting in typical sigmoidal psychometric curves (Fig. 4.3 top rows). Second, the learning curves followed the exponential decay function found in all PL paradigms (Fig. 4.3 bottom rows).

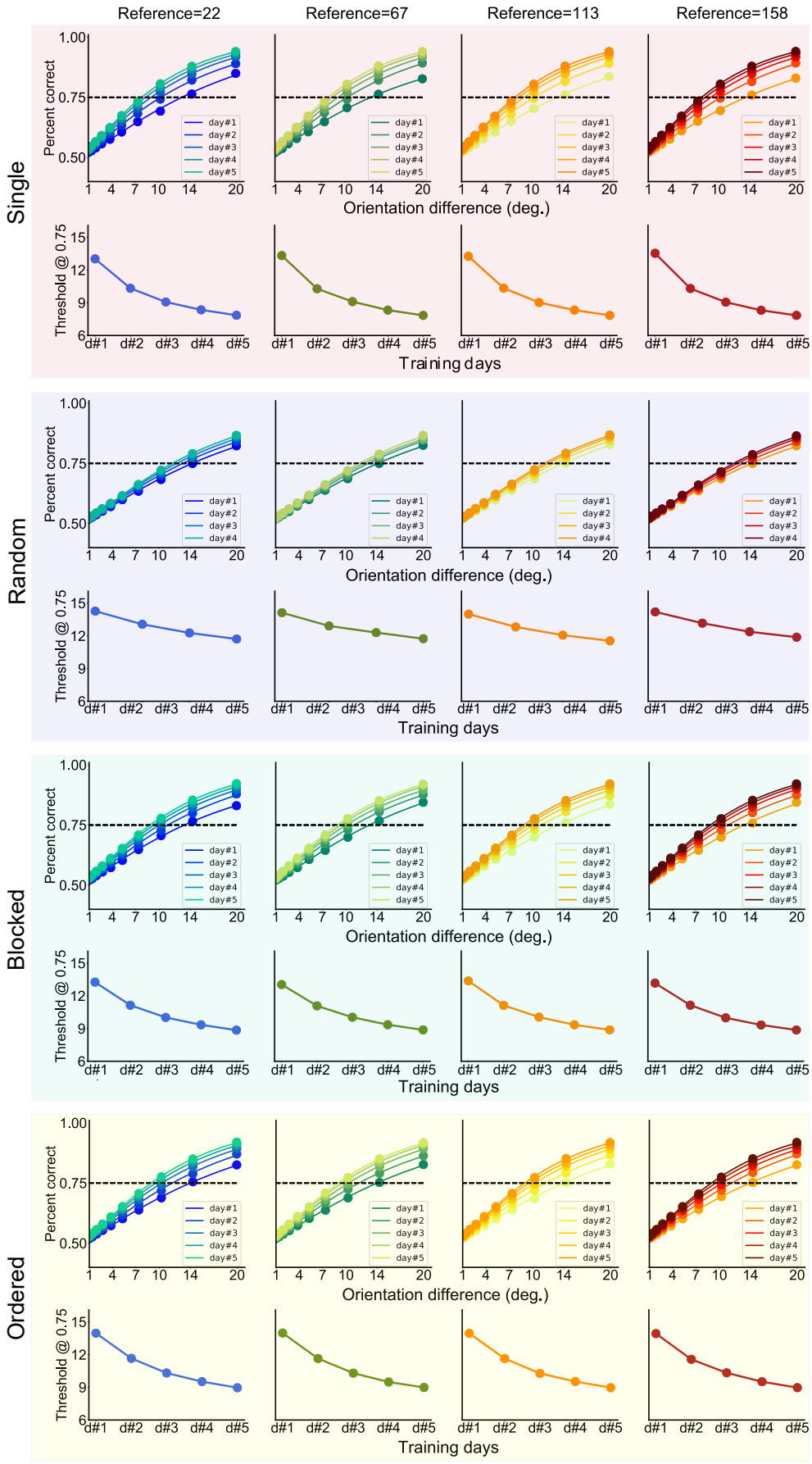


Figure 4.3: The psychometric & learning curves in the four simulated PL experiments. Caption continues on the next page.

Figure 4.3: (Caption for Fig. 4.3 on the previous page.) **Top rows in all subpanels:** The psychometric curves show the percent correct values (y axis) as a function of the absolute increments in degrees (x axis) for each training day (colored dots). The colored lines are fitted sigmoidal functions to the percent correct values. Dotted lines show the 75% correct performance level. **Bottom rows in all subpanels:** The learning curves show the discrimination thresholds at 75% correct performance level (y axis) as a function of the training days (x axis). **All panels:** each column represents the discrimination performance from the reference value shown on the top of the figure. Note that in the single experiment the columns corresponding to the different reference values show separate simulations for a 5-day training with a single reference value.

To compare the performance of the BSPL model to the behavioural results found in previous studies, two parameters were fitted to match the initial performances in orientation discrimination tasks and the amount of learning found in the randomly interleaved roving experiments. First, the amount of resources (see N in Eqs. 4.24 and 4.25), modulating the noise in the sensory observations through the Fisher information (see Eq. 4.7), was set in the single reference experiment to loosely match the average initial thresholds that we found in our experiments using orientation discrimination tasks (Lengyel & Fiser, 2019). All other experiments used this resource constraint during the response probability generation. Second, the learning rate (see α in Eq. 4.29), determining how much the current observations influence the prior resource allocations, was fitted in the randomly interleaved experiment to the average of the average amount of learning found in the previous experiments in the randomly interleaved roving condition (see the second bar in Fig. 4.4, left bar chart). The other experiments, then, used this fitted learning rate.

The BSPL model replicated the pattern of results found in classical and roving studies (Fig. 4.4, bar chart on the left); the amount of learning was the smallest in the randomly interleaved experiment while most learning appeared in the classical experiment with one reference (Fig. 4.4, bar chart on the right). Crucially, substantial learning took place in the blocked and in the fixed order conditions demonstrating the influence of SL on PL. These results are in line with the previous findings (compare left and right bar charts in Fig. 4.4, also see Section 4.2.1).

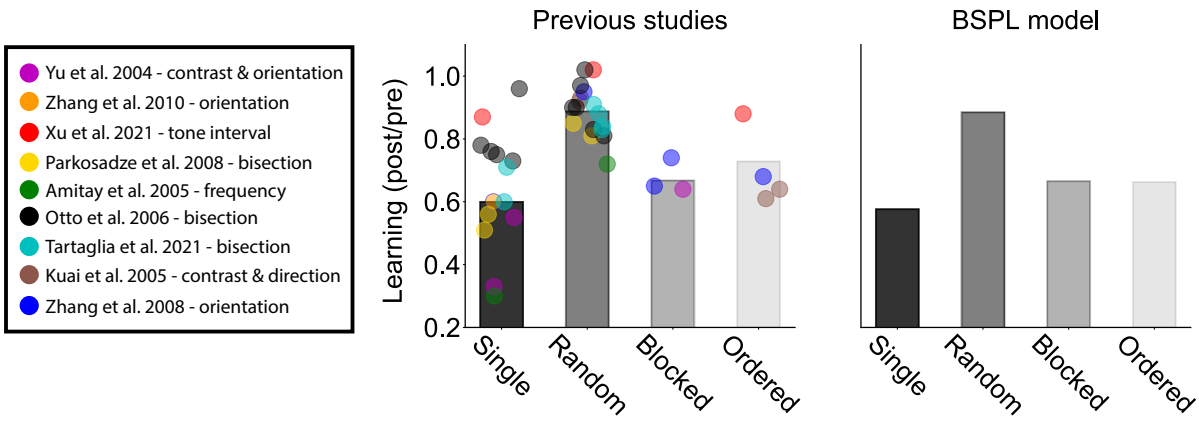


Figure 4.4: Comparing the amount of learning in the BSPL model and in previous studies. Right: The amount of learning, quantified as post/pre thresholds, in the four simulated experiments using the BSPL model. The thresholds in the four reference-contexts were averaged. **Middle:** The average of the average amount of learning found in previous studies in the four types of roving conditions. The amount of learning was defined in the same way (post/pre) and colored dots represent the average learning across participants in separate experiments. Dots in the same color represent experiments conducted in the same study. **Left:** The references of the studies that appear on the bar chart in the middle and the perceptual attributes used in those experiments. Post/pre represents the thresholds in the last day (see Fig. 4.4, d#1) divided by the thresholds in the first day of practice (see Fig. 4.4, d#5).

Similar to the findings of Parkosadze et al. (2008), demonstrating improvement after a longer, 10-day long practice in the randomly interleaved condition, the simulation results for longer training in the BSPL model also show substantial amount of learning in the randomly interleaved condition (Fig. 4.5).

These results demonstrates that inferring the reference-context in the trials and learning the transition model between the reference-contexts during SL can support PL modeled as the adaptation of the neural resources optimizing the encoding of the stimuli.

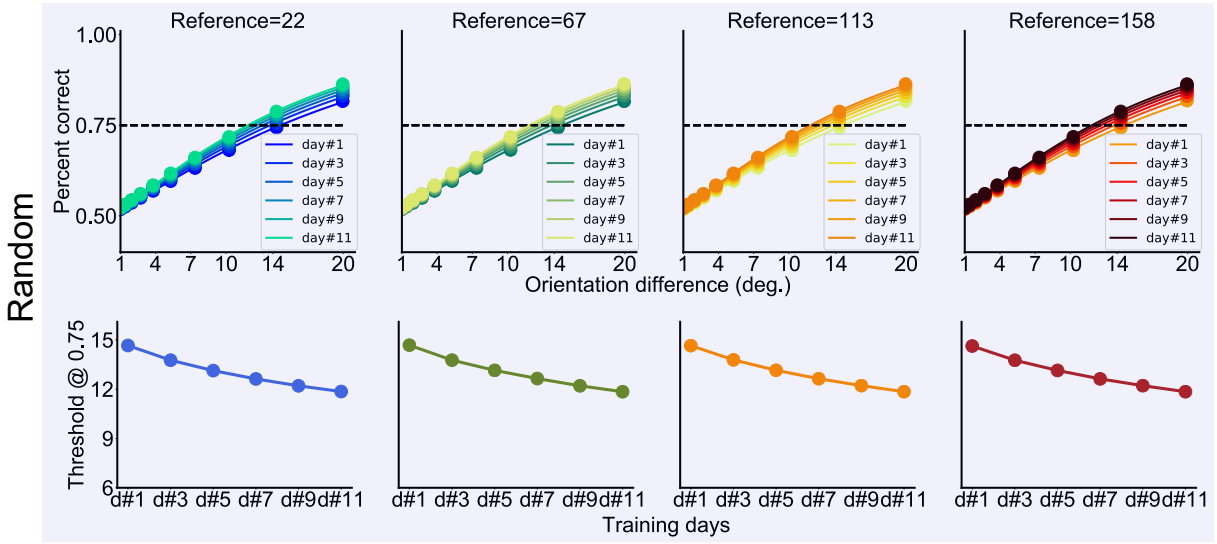


Figure 4.5: The psychometric & learning curves in the randomly interleaved experiment with long training. The average amount of learning, measured as post/pre thresholds, was 0.8 in the longer training compared to the 0.9 in the shorter training (see Fig. 4.3, blue subpanel and Fig. 4.4, right bar chart). See the caption for Fig. 4.3 for more information.

4.5 Contrasting the BSPL with other PL models

PL has been associated with a wide range of neural learning effects such as response gain modulation, tuning curve sharpening, shifts in the neurons' preferred feature values, reduced variability and correlations, and refined routing and connections (see the details and the references in Section 1.1.2). As a result, (1) several learning mechanisms were proposed to capture the behaviour during PL tasks (e.g., see Dosher and Lu, 2017; LeMessurier and Feldman, 2018; Schwabe, 2005; Teich and Qian, 2003) and (2) most successful PL models did not specified representations and learning processes (e.g., Ahissar and Hochstein, 2004; Shibata et al., 2014; Watanabe and Sasaki, 2015) in detail so that they could flexibly explain the behaviour and neural correlates of PL. Next, I will compare the BSPL model to the three most influential computational models in PL: the reverse hierarchy (Ahissar & Hochstein, 1997, 2004), the two-stage (Shibata et al., 2014; Watanabe & Sasaki, 2015), and the reweighting (Dosher & Lu, 2017; Dosher et al., 2013) models.

First, I address the two-stage and the reverse hierarchy models together because neither of them specifies the latent representations or the learning mechanism underlying PL. Both frameworks emphasize the need to consider the interaction between learning the structure of the task and stimuli and adapting the perception of the stimuli to the learnt structures. The two-stage model states that there are perceptual feature- and cognitive task-based plasticity during PL (Watanabe and Sasaki, 2015) while the reverse hierarchy theory posits that learning follows a gradual top-down direction which starts at the highest, cognitive level and only later, after extensive training, expands to lower-level perceptual processes (see Ahissar and Hochstein, 2004). Conceptually, the BSPL model is very similar to these models and relies on the interaction between learning the structure in the stimuli and adapting perception to the learnt structures. However, in contrast to these models, the BSPL model provides a normative modelling framework that in influence of the the top-down influences of the higher-level, abstract latent variables, capturing the task and stimuli structures, on the lower-level perceptual processes.

Second, the reweighting framework (Dosher & Lu, 2017), the most successful computational framework in PL, assumes that learning reflects the improvement in decoding (or reading out from) the sensory representation instead of assuming an improvement in the sensory encoding of the stimulus. Most reweighting models posit that the neural activity of the sensory neurons, encoding the feature of the stimulus, are pooled with different weights when decoding the value of the stimulus's feature or the trial's decision category and during learning, the pooling weights are being adjusted to improve the decoding performance (Dosher et al., 2013). These feed-forward neural networks in the reweighting models can be considered as recognition models of the stimulus's feature value or the trial's decision category. Therefore, one can find the equivalent generative models for the recognition models in the reweighting framework and compare the BSPL model directly to those equivalent generative models. It will be for future

studies to derive the generative models corresponding to the reweighting models and compare them to the BSPL model under the same probabilistic framework.

4.6 Predictions & future directions of the BSPL model

This chapter only represents the first step to develop a unifying modeling framework for PL & SL and future work, among others, can investigate the novel predictions of the BSPL model in new experiments, perform model fitting and selection on experimental data, and test other modeling choices and assumptions in the BSPL model. In this section, I only describe two predictions regarding generalization of learning.

First, the BSPL model predicts that the observer can generalize the learned transition model between reference-contexts to other task and to somewhat altered stimuli. A testing scenario of this generalization would be that after practicing discrimination for several days with one perceptual attribute in the fixed order roving condition, the perceptual attribute in focus changes in the roving task, but the original perceptual attributes of the reference-stimuli remains visible keeping the temporal structure between the reference-contexts the same. The initial threshold with the second perceptual attribute is expected to be smaller when the structure in the reference-contexts is transferred from the previous training with the first perceptual attribute compared to when no prior knowledge is assumed on the structure in the reference-contexts. Fig. 4.6 demonstrates this generalization study by showing the synthetic performance levels in two experiments with fixed order roving conditions using contrast discrimination. In the first experiment (Fig. 4.6, top subpanel) the transition model of the reference-contexts was inherited from an orientation discrimination task with the fixed order roving condition after five days of training. In the second, control experiment (Fig. 4.6, bottom subpanel), the transition model of the reference-contexts had a uniform prior. The initial threshold were slightly higher, and the amount of

learning was a little bit larger in the experiment in which the transition model was generalized from the previous training with another perceptual attribute than the initial thresholds and learning in the control experiment (see the caption of Fig. 4.6 for more detail).

Second, previous studies found that, after PL emerged in the blocked and fixed temporal order conditions in roving, learning generalized entirely to other temporal structures and to random reference patterns (Kuai et al., 2005; Zhang et al., 2008). This transfer of learning is not predicted from the BSPL model, however there are two ways to account for this powerful generalization. First, the learning rate (see α in Eq. 4.29), determining how much the current observations influence the prior resource allocations, could increase in the beginning of every new blocks, or when there is a large discrepancy between the current observations and the prior expectations learned based on past observations increasing the uncertainty in the inference. Several previous studies found that the learning rate fluctuates in an optimal way; in volatile environments, in which inference has larger uncertainty, the learning rate is higher than in stable environments, when participants are able to make more certain inferences (e.g. Behrens et al., 2007; Piray and Daw, 2020). An other way in which the BSPL model could produce full generalization to randomly interleaved references is if the participant after observing the reference-stimuli reallocate the resources incorporating the information from the observation (see the last paragraph in Section 4.3.4). Then, the resource allocation will be optimized for the right reference-context when encoding the test-stimulus resulting in better discrimination performance too.

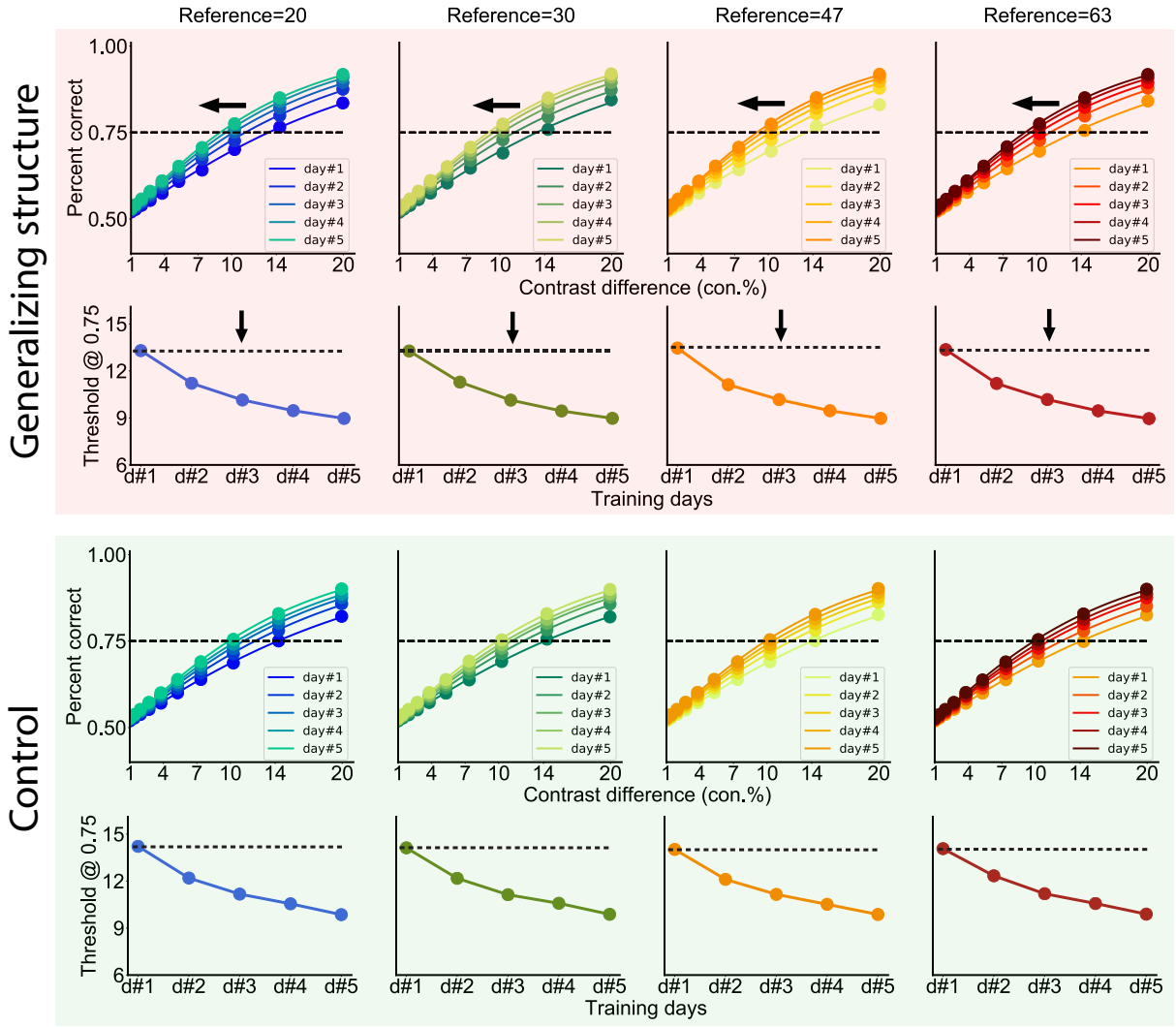


Figure 4.6: Generalizing the structure between the reference-contexts to PL tasks with other perceptual attributes. The psychometric (top rows) & learning curves (bottom rows) in two fixed order roving experiments using contrast discrimination. **Top, red subpanel:** The synthetic observer transfers the transition model between the reference-contexts from a previous experiment. **Bottom, green subpanel:** the observer had no prior information about the transition model between the reference-contexts. The initial thresholds were slightly lower in the generalization experiment (13 con. %) compared the control experiment (14 con. %) with uniform prior over the transition model between the reference-contexts. The average amount of learning, measured as post/pre thresholds was also smaller in the generalization experiment indicating more learning ($\frac{post}{pre}_{\text{Generalization}}$ =0.66), than in the control experiment ($\frac{post}{pre}_{\text{control}}$ =0.69). The dotted lines on the psychometric curves marks the 75% correct performance while on the learning curves they show the initial threshold levels. The arrows on the top subpanels highlight the small differences between the amounts of learning (top row) and the initial thresholds (bottom row) in the generalization and the control experiments.

4.7 Conclusion

The BSPL model provides a unifying normative framework for both classical and roving paradigms in PL. Since roving paradigms inherently combines PL & SL, the BSPL model is suitable for exploring the interaction between the two learning types in simple perceptual tasks. The key characteristics of the BSPL model is that it infers the contexts of the trials and learns the temporal transition model between the trial's contexts during SL. This SL process, then, interacts with the PL process in two ways. First, the prior probability over the trial's context can influence the probability of the stimulus values in the trials. Second and more importantly, inferring the trial's context determines the neural resource allocation for efficient stimuli encoding (which drives PL) in the contexts. This interaction, in the BSPL model, can account for the wide range of findings in PL using roving and provides testable predictions for new experiments.

Chapter 5

General Discussion

In this thesis, I proposed a unified framework for perceptual (PL) and statistical learning (SL) that can seamlessly integrate recent findings showing interactions and shared computational principles between the two learning types in behavioural experiments, and overlapping neural correlates in imaging and neurophysiological studies (Chapter 1). In this framework, computations of the two forms of learning are captured within the same Hierarchical Bayesian Model (HBM), and the probability distributions in the HBM are assumed to be represented with a sampling-based neural coding in the brain.

Using this framework, first, I presented an empirical study investigating two previously found, general rules predicting learning and generalization performance in classical perceptual learning (Chapter 2). We confirmed the first rule that posited a Weber-like relationship between initial performance and the amount of learning. However, we also showed that the Weber-like relationship does not reflect any general characteristics of the learning in PL tasks, it only shows how the observer relates physical intensities to perceived magnitudes during perception. Second, we found that the amount of generalization was proportional to the amount of learning. Together with previous findings demonstrating that more training examples elicit less general-

ization, this result suggests that variability and the number of repetitions of the stimuli during the practice together influence the amount of transfer in learning and that there is no stimulus- and task-independent general rule that can predict generalization in PL paradigms.

In the next two studies, I investigated how learning the structure in the stimuli (i.e., SL) influences perception by focusing on the formation of object representation (Chapter 3). First, using a series of behavior experiments, I showed that statistically defined chunks learnt within the classical spatial SL paradigm elicit perceptual and attentional processes very similar to what real visual objects do. Second, in a study, combining visual and haptic stimulation, I demonstrated that participants instantaneously build abstract, amodal representations of the chunks defined either by visual or by haptic statistical properties alone in a setup that allowed only zero-shot across-modality generalization. These results together suggest that the processes guiding SL lay at the very heart of one of the most fundamental aspect of perception, object representation, as they influences how observers segment their sensory input into perceptual units. Furthermore, the attentional and generalization effects linked to SL-based chunk representations in this chapter are indicative of hierarchical probabilistic computations in the brain that are in line with the Bayesian latent variable model defined in the HBM framework for interpreting SL.

Finally, I formally investigated the interaction between SL and PL based on the extensive results reported on PL using the roving paradigm by developing a unifying Bayesian model that can explain behaviour in both classical and roving types of PL experiments (Chapter 4). The model assumes that the observer interprets the experimental setup in a hierarchical manner by implicitly detecting and mentally representing reference-contexts across the trials and gradually learns the transition model between the reference-contexts via SL if the transitions have a noticeable structure. In return, the obtained transition model supports an efficient resource allocation for stimulus encoding in the reference-contexts allowing a successful PL of

the discrimination task. This formally captured interaction between PL and SL in the Bayesian statistical perceptual learning model can explain a wide range of behaviour results found in the PL literature using roving and beyond.

The findings in the three empirical and in the final simulation studies presented in this thesis have broad implications for sensory learning by supporting the proposed unifying framework for PL and SL, and for object perception by backing the idea that "objectness" can parsimoniously be explained by sufficient sets of statistical contingencies. In the next sections, I will first discuss these implications and then address future directions of this line of research and the novel predictions that the unifying framework of sensory learning offers.

5.1 Sensory learning

Investigating sensory learning is essential to understand perception since perception hardly exists in complete isolation from learning processes. Naturalistic scenarios contain rich contexts and complex tasks, therefore, when studying perception researchers need to address the following two learning processes jointly: (1) Learning the statistical structures of the features, the contexts, and the tasks (SL), and (2) the adaptation of perception in the light of the learnt statistical properties of the features, contexts, and tasks (PL).

These two domains in sensory learning have been treated in complete separation which seemed reasonable given the large differences in the methods and results of the traditional PL and SL paradigms (see Sections 1.1 and 1.2 for details). However, in this thesis, I argued that the two forms of learning should be investigated under one, unifying sensory learning framework that could capture both the classical findings that originally fueled the arguments for separating PL & SL, and more recent studies, that demonstrated overlapping mechanisms between the two learning types.

In contrast to earlier PL studies, new studies showed strong and context dependent transfer of learning in a wide range of different PL paradigms (Chang et al., 2013; Chang et al., 2014; Green et al., 2015; Green et al., 2010a; Kattner et al., 2017; Kuai et al., 2005; Wang et al., 2016; Wang et al., 2014; Xiao et al., 2008). Studies, using imaging and neurophysiological techniques, also demonstrated that not just the low-level sensory areas, but a large set of brain regions, including high-level, cognitive and decision-making areas, are active during the learning process in classical PL paradigms (Diaz et al., 2017; Kahnt et al., 2011; Law & Gold, 2008, 2010; Maniglia & Seitz, 2018). These results suggest that (1) there are strong top-down modulations during PL, and (2) that not just low-level, sensory, but more abstract, higher-level representations are involved in the PL process.

Roving paradigms provide an especially direct evidence that the statistical structure of the stimuli modulates the extent of learning in classical PL tasks. The results of these roving studies cannot be parsimoniously explained without the consideration of an SL process during which the observers learns the temporal structure between the references which, then, interacts with and supports PL. The first empirical study, presented in the present thesis (Section 2), further supports this claim: I showed that in contrast to previous reports suggesting common laws of learning and generalization in PL, the amount of learning and transfer are highly depended on the statistical structure of the stimuli and the task and, therefore, SL is needed to be integrated with PL to successfully predict learning and generalization performances in PL paradigms.

I argued that existing modelling frameworks of PL could not directly address the above listed phenomena showing context dependent generalization, the involvement of higher-level brain areas, and that learning is influenced by the statistical structure of the stimuli and the task. In particular, the reweighting framework (Dosher & Lu, 2017), does not incorporate computation-

ally any top-down modulations of stimuli and task structures, consequently it cannot address the influence of context and stimuli structure in PL tasks. This is not to say that convergence between reweighting and HBN models is impossible, since the generative model of the HBM framework itself can be defined in multiple ways and each generative model can be paired with a number of recognition models including ones that are similar to models of the reweighting framework. However, key characteristics of the HBN framework, the representation of uncertainty at multiple levels, the bi-directional interaction in the inference process and the extended scope of the sensory structure that learning must capture are crucial to maintain the power of the derived model to capture human behavior.

Despite the fact that both two-stage models (Shibata et al., 2014) and the Reverse Hierarchy Theory (Ahissar & Hochstein, 2004) assume top-down influences that can explain the effect of the stimuli and task structures, there are large differences between these models and the HBM framework proposed here as well. First, the HBM is a fully Bayesian framework, i. e. it represents the uncertainty of all its latent variables. While in a given recognition model the requirement for a fully-Bayesian representation might be partially relaxed, a substantial part of the model must retain uncertainty representation to remain capable of strong generalization (Koblinger et al., 2021). Neither the two-stage models not RHT discuss the treatment of uncertainty, and integrating the concept would shape those frameworks to be more similar to HBMs. The second difference is that HBMs are normative, i.e. they incorporate the top-down influences of the higher-level, abstract latent variables by calculating the statistically optimal combination of information at the different levels of the hierarchical structure defined among the latent variables. While this feature can be incorporated into the other two frameworks, such a computational treatment has not been discussed even though it is also essential for capturing the flexibility and characteristics of human behavior (Koblinger et al., 2021).

I also showed in my thesis that significant interactions between PL and SL can be detected not only in PL but in SL paradigms as well through modulations measured in a wide range of perceptual processes including image detection, numerosity and size perception (Barakat et al., 2013; Luo & Zhao, 2018; Otsuka & Saiki, 2016; Piazza et al., 2018; Sotiropoulos et al., 2011; Yu & Zhao, 2018; Zhao et al., 2013; Zhao et al., 2011; Zhao & Yu, 2016), and that, similarly to PL, the neural correlates of SL can also be found at the lowest level of cortical representations (Karlaftis et al., 2018; Köver et al., 2013; Wang et al., 2017) . These findings reinforce the idea that in order to describe learning in SL paradigms, one has to characterize how the perception of relevant features adapts to the statistical structure of the stimuli (PL) as well, beyond describing how the statistical structures themselves are learnt (SL). The two empirical studies, investigating SL in this thesis, provide further support to this claim by demonstrating that one of the most fundamental aspects of perception is largely determined by SL; the way how the sensory input is segmented into meaningful units (see Chapter 3).

So far, computational models in SL have ignored the low-level perceptual processes and focused only on how the structure in the stimuli is learnt (Mareschal & French, 2017; Orbán et al., 2008; Perruchet, 2019). Although some past work took lower-level perceptual features into consideration in their models (Austerweil & Griffiths, 2011; Froyen et al., 2015; Yildirim & Jacobs, 2012, 2013), they did not incorporate perceptual learning processes. The unifying HBM framework treats PL and SL jointly to better capture the learning results obtained with both learning paradigms and thus, it could set up the path for a fuller understanding of sensory learning in naturalistic scenarios.

To sum up, the results in the PL and SL literature imply that the two forms of learning

share characteristics in almost every domain and that none of the two learning types can be investigated properly without considering the other learning type. The strict separation between SL and PL is only meaningful if one focuses on the extreme testing paradigms of PL and SL and disregards all other more complex and naturalistic scenarios. Outside of the lab, naturalistic stimuli and tasks contain rich structures and when adapting to those structures during sensory learning, PL and SL processes operates jointly to track and represent the statistical regularities in the environment and to adapt the perception of features to the regularities.

5.2 Object perception

One of the most fundamental aspect of perception is that our brain organizes the incoming sensory information into distinct, meaningful units, called objects. In the two empirical studies in Chapter 3, we investigated the "objectness" of statistically defined chunks during classical SL tasks to asses the role of SL in the emergence of object representations. First, we found that chunks solely defined by consistent statistical properties engaged classical object-based attention and object-based perceptual processing. Second, we demonstrated zero-shot generalization effects between visual and haptic modalities after learning statistical regularities only in one of the modalities. Based on these findings, I propose that observers create abstract, amodal representations of chunks, defined by consistent statistical properties, that serve as perceptual units for subsequent sensory processing in a similar way to how the representations of real objects form the units in perceptual organization.

This proposal implies that objects should be defined as a sufficient set of statistical contingencies and objectness should be treated as a continuum rather than an all-or-none feature. In this proposal, the degree of objectness depends on the level of statistical coherence; the stronger

the coherence is the more object-like the representation is. In natural scenes, real objects are indeed a particularly strong aggregation of statistical contingencies. Considering the classically defined object features, such as long contours, similar textures/colors, and Gestalt structures, they are all exemplars of strong statistical correlations. Furthermore, there are several examples when objectness is ambiguous due to less coherent statistical patterns of the object feature, e.g., a jet of steam, or when ambiguity is caused by irrelevant statistical cues, e.g., the illusory contours of the Kanizsa triangle, or any example of animal camouflage.

What cues are necessary and/or sufficient in general to identify some sensory input as an object is a subject of intensive debate in the literature (Kellman & Spelke, 1983; Palmer & Rock, 1994; Spelke, 1990). According to previous studies, stable boundaries such as luminance contours are the strongest criteria for objectness (Kellman & Spelke, 1983; Palmer & Rock, 1994; Spelke, 1990), and this is the reason why studies investigating object-based perceptual effects typically used clear luminance contours to define objects (Egly et al., 1994; Lee et al., 2012; Moore et al., 1998; Shomstein & Yantis, 2004). However, such contours are nothing else but strong statistical contingencies of luminance edge segments appearing and moving in statistical coherence. Moreover, there are several studies suggesting that having stable boundaries based on luminance discontinuities is not a necessary criterion for "objectness" as objects can be defined by texture-, disparity-, regularity-, symmetry-, or motion-based “boundaries” (Feldman, 1997; Julesz, 1971; Schofield, 2000). Similarly, in natural scenes there are many potential luminance contours that are inside objects and thus do not correspond to actual physical boundaries separating objects. Thus, relying solely on luminance contours to segment the environment into objects would lead to many errors, and it is reasonable to assume that there is a more general principle underlying perceptual organization: consistent statistical properties.

Real mental objects emerge over years of interaction with the environment and the full richness of those mental constructs are attained by learning a large set of consistent statistical properties across many scenes, contexts, and tasks. It will be for future studies to investigate whether and how statistical learning mechanisms can by itself produce real mental objects or what additional mechanisms are needed. Future work can also address the relative importance of statistically learned and innately available cues and representations in the development of object representations. In the two studies in Chapter 3 only the first steps were made by claiming that conceptually, there is a close link between chunking based on statistical learning and object segmentation based on visual boundary cues: both kinds of segmentation rely on statistical coherence, they both serve as fundamental components in forming object representations, and they are both sufficient to elicit some object-based perceptual effects. These results may encourage the field of developmental cognitive science to refine the definition of an object, study the necessary and sufficient conditions for object segmentation, and investigate the role of statistically learned and innately available cues in the emergence of object representations.

5.3 Future directions

The integrated viewpoint of sensory learning, proposed in this thesis, provide useful guiding principles to design future experiments that can bring PL and SL paradigms closer to each other in a systematic way. Experiments that involve both types of learning paradigms will contain more complex stimuli and richer statistical structures than the stimuli in classical PL paradigms. Furthermore, in contrast to classical SL paradigms, a joint PL-SL paradigm will always have a task that could assess the performance online. These characteristics of the new PL-SL paradigms will shift future studies to more naturalistic stimulation consisting of complex structures, which structures will be utilized to achieve a better performance in the task at hand.

The only presently available example in the literature for a joint PL-SL paradigm with extensive behavioral results is roving. Previous reports have already demonstrated that regularities in the intermixed properties of the stimuli influence the amount of learning in fine discrimination task (Cong & Zhang, 2014; Kuai et al., 2005; Zhang et al., 2008). This line of research can be extended to investigate spatio-temporal regularities with different complexity. Based on the paradigm suggested in Fig. 1.6C in Chapter 1 future research can study different order of temporal correlations in the reference-stimuli and introduce spatial co-occurrence statistics between the stimuli along with the temporal regularities.

Another strategy to attain a fuller understanding of sensory learning is to design experiments with naturalistic stimuli and tasks, in which both forms of learning are inherently important. A recent paradigm attracting researchers studying motion perception in the past few years could provide an ideal testing scenario with a naturalistic task and moderately complex stimuli. In this paradigm, the participant has to catch a target in a dynamic environment (Kwon et al., 2020b; Lakshminarasimhan et al., 2018). In this "catching" task, the participant receives sensory inputs suggesting both self and object motions, while the environment contains background textures as well as irrelevant and target objects. This paradigm inherently involves PL since the participant has to fine-tune the perception of motion direction and velocity, texture discrimination, and object detection to achieve better performance in the task. SL processes are also salient in the paradigm since the participant has to learn and adapt to the statistical regularities in the locations, trajectories, and velocities of the objects. This paradigm can also be extended to include statistical structures in the background textures providing contextual information or by using multiple target objects with spatio-temporal structures.

The unifying HBM framework provides a suitable scheme to explain existing results in rov-

ing studies (Adini et al., 2002; Adini et al., 2004; Amitay et al., 2005; Banai et al., 2010; Cong & Zhang, 2014; Dosher et al., 2020; Kuai et al., 2005; Nahum et al., 2010; Otto et al., 2006; Parkosadze et al., 2008; Tartaglia et al., 2009b; Yu et al., 2004; Zhang et al., 2008), double-training paradigms (Wang et al., 2014; Xiao et al., 2008; Xiong et al., 2016), imagination-based learning (Tartaglia et al., 2009c), experiments connecting categorization and PL (Tan et al., 2019), and experiments demonstrating the modulation of perceptual processes due to SL (Luo & Zhao, 2018; Otsuka & Saiki, 2016; Piazza et al., 2018; Yu & Zhao, 2018; Zhao et al., 2013; Zhao et al., 2011; Zhao & Yu, 2016).

The HBM capturing roving results combines a possible PL mechanism, that adapts to improve performance in the task, with a hidden Markov model that tracks the regularities in the stimuli. Using this HBM model, the existing behavioural results can parsimoniously be explained with the interaction between SL and PL. Importantly, the HBM offers several new predictions for future studies using similar roving paradigms.

First, the HBM can predict the performance in roving experiments using different structure in the reference-stimuli. For example, when the references would form pairs and the elements in the pairs would appear in a particular consecutive order during the task (see Fig. 1.6C), the HBM model predicted better performance for the references which constitute the second elements in the pairs because those references have higher predictive probabilities after observing the first element from the reference-pair than the predictive probabilities of the other references which, being the first elements in the pairs, can appear after any of the second element references. Another prediction of the HBM is that the learnt structure in the stimuli should generalize to other perceptual tasks. Therefore, in roving paradigms, after training with references, following a fixed temporal order, if the relevant perceptual attribute, based on which the discrimination was performed, would change, but the stimuli and its pattern would remain the

same the HBM predicted a substantial amount of improvement in the performance due to the learnt, thus predictable stimuli statistics (Fig. 4.6).

Regarding the modelling of the more complex naturalistic "catching" PL-SL task mentioned above, previous studies have already implemented HBMs to explain the computations how participants solve the catching task in such dynamic environments (Kwon et al., 2020b; Wu et al., 2018a, 2018b). However, applying HBMs for dynamic scenes in general results in intractable computations and algorithms implementing approximate Bayesian inference in such scenarios are still scarce (but see Ellis et al., 2018; Kwon et al., 2020b). This problem of computational complexity has to be solved for introducing a PL mechanism in the HBM. The joint modelling of PL & SL raises the problem of connecting the abstract, computational level modelling of SL with the lower-level, neural, implementational level descriptions of PL. Using HBMs seems to further complicates this connection, but the level of understanding of how the brain might implement Bayesian inference has been steadily increasing in the past decades and presently, there are several successful implementational frameworks for HBMs in the neuroscience literature. E.g., the probabilistic population codes framework assumes that the firing activity of the neurons represents parameters of the posterior distribution in a logarithmic space (Ma et al., 2006). Extended models in this framework successfully formalized implementations for probabilistic computations in HBMs (Deneve, 2005; Vasudeva Raju & Pitkow, 2016). However, these models are limited to using marginal posterior distributions and cannot handle joint posteriors over multiple latent variables. In an alternative sampling-based, direct variable coding framework, joint posteriors can be captured by assuming that the neural activity over some time period directly represents a sampling-based approximation of the joint posterior distribution over the latent variables in a HBM (Echeveste et al., 2020; Fiser et al., 2010; Orbán et al., 2016). Researchers have provided substantial neural evidence supporting the sampling-based framework

by analyzing the static (Haefner et al., 2016; Orbán et al., 2016) and dynamic (Echeveste et al., 2020) activity patterns of the primary visual cortex.

Most implementational frameworks aim at capturing Bayesian inference and not Bayesian learning. Although the computational principles of inference and learning are similar, the two can diverge at the implementational level. For example, it is not straightforward whether and how the brain can implement probabilistic learning by representing the uncertainty in the unknown and continuously adapted parameters although interesting proposals exist about how synapses might take their own uncertainty into account (Aitchison et al., 2021).

Introducing a PL mechanism in any of the implementational frameworks raises further complications. Several neural mechanisms were found to be correlated with PL, among others, improved neural encoding with adapted tuning curves, reduced variability and correlations, refined routing, connections, and circuit-level dynamics were proposed to induce PL (Dosher & Lu, 2017; LeMessurier & Feldman, 2018; Schwabe, 2005; Teich & Qian, 2003). Formalizing and exploring these mechanisms in a HBM using any of the implementation frameworks to explain behavioral and neural data in the PL literature posits great challenges for future modelling works.

The application of the unifying HBM framework in experiments combining PL and SL paradigms poses several challenges of identifying the neural mechanisms of PL and applying the mechanism in a HBM. The space of existing models in the PL literature has been substantially narrowed down to models related to the reweighting framework, but there is a need for new frameworks that can address more complex paradigms in PL than the classical ones, and developing HBMs that can connect PL & SL is one step to develop such a unifying framework.

5.4 Conclusions

The integrated framework proposed in this thesis, that unifies perceptual (PL) and statistical learning (SL) with hierarchical Bayesian modelling (HBM), provides a parsimonious explanation for the diverse set of previous results suggesting shared underlying processes between PL and SL. This common probabilistic approach offers suggestions for new types of experiments that would shift investigations in sensory learning towards more naturalistic simulations. The joint modelling of PL and SL will require researchers to connect abstract probabilistic models to learning mechanisms at the neural level bringing computational and implementational level modelling closer to each other. This unified framework has encouraged the investigations of two previously found general laws predicting learning and generalization in PL (2), and the influence of SL on segmenting the sensory input into objects (3). In these studies, first, we found that there are no general laws that can predict learning and generalization in PL paradigms and second, we demonstrated that SL has a pivotal rule in object perception. Importantly, the joint treatment of PL and SL can also encourage future studies to investigate the predictions of using a common HBM in complex learning situations such as roving experiments (4), recent SL studies, and more complex naturalistic paradigms. Real life stimulation contains rich structures and complex tasks. To adapt to those regularities during sensory learning, humans need to track and represent the statistical correlations in the environment and adapt their perception to those regularities through a sophisticated machinery. The present thesis lay down some essential requirements of defining such a machinery in a computationally and biologically feasible way.

Appendices

Appendix A

Supplementary materials for study 1

A.1 Quantifying the decrease in thresholds and lapse rates

Estimating discrimination thresholds using adaptive staircase methods confounds errors due to lapses (lack of attention) with errors due to real perceptual indiscriminabilty (Solomon & Tyler, 2017). Although the 3-down-1-up staircase can be robust to the initial attentional lapses (Karmali et al., 2016) lapses are not necessarily limited to the initial trials in novice observers. We investigated this potential confound by first estimating the lapse rates and the discrimination thresholds for each observer in the pre- and posttests by fitting psychometric curves to the observers' performance and then, testing whether the thresholds and/or the lapse rates decreased due to learning. We confirmed that both participants' thresholds and attentional lapses decreased due to practice.

We fitted cumulative Weibull distributions (psychometric curves) to participants' data at the pre- and the posttests:

$$P(x) = \varepsilon + (1 - \varepsilon - \gamma)(1 - e^{-(\frac{x}{\alpha})^{\beta}}) \tag{A.1}$$

In this formula, x is the stimulus strength which is the contrast and the orientation difference in

% contrast and in degrees respectively. $P(x)$ is the fraction of the correct responses at stimulus strength x. The parameters denote the following: α is the threshold, β is the slope, γ is the lapse rate, and ε is the chance performance level which was 0.5 in our tasks. The lapse rates and the thresholds of the observers were estimated by the best-fitting value of the γ and α parameters using maximum likelihood estimation.

The lapse rates decreased significantly due to training in the contrast discrimination with within-subject design (t_{16}=3.154, P=0.006, d=0.786, Fig. A.1, top row, in the middle) and in the orientation discrimination experiments (t_{29}=3.226, P=0.003, d=0.599, Fig. A.1, top row, on the right), however it did not change in the contrast experiment with between-subject design (t_{30}=0.282, P=0.779, d=0.052, Fig. A.1, top row, on the left).

The thresholds of the participants decreased significantly in all experimental condition after training. In the orientation discrimination experiment we obtained t_{14}=5.834, P<0.001, d=1.559 with reference orientation 0°, and t_{14}=3.319, P=0.005, d=0.882 with the reference orientation 25° (Fig. A.1, middle row, on the right). In the contrast discrimination experiment with between-subject design we obtained t_{14}=2.969, P=0.010, d=0.793, at reference contrast 30%, and t_{15}=3.298, P=0.005, d=0.851 at reference contrast 73% (Fig. A.1, middle row, on the left). In the contrast discrimination experiment with within-subject design we obtained t_{15}=3.137, P=0.007, d=0.809, at reference contrast 30%, and t_{15}=3.748, P=0.002, d=0.968 at reference contrast 73% (Fig. A.1, middle row, in the middle).

We compared the two measurements assessing the decrease in the observers' thresholds due training. In all experiments there were large positive correlations between the decrease in thresholds estimated by the staircase and by the best-fitted Weibull function. These correlations were r=0.66, P<0.001, CI_{95}=0.40-0.83 in the contrast experiment with between-subject design (Fig. A.1, bottom row, on the left), r=0.69, P<0.001, CI_{95}=0.45-0.84 in the contrast experi-

ment with within-subject design (Fig. A.1, bottom row, in the middle), and r=0.79, P<0.001, CI_{95}=0.59-0.90 in the orientation experiment with between-subject design (Fig. A.1, bottom row, on the right).

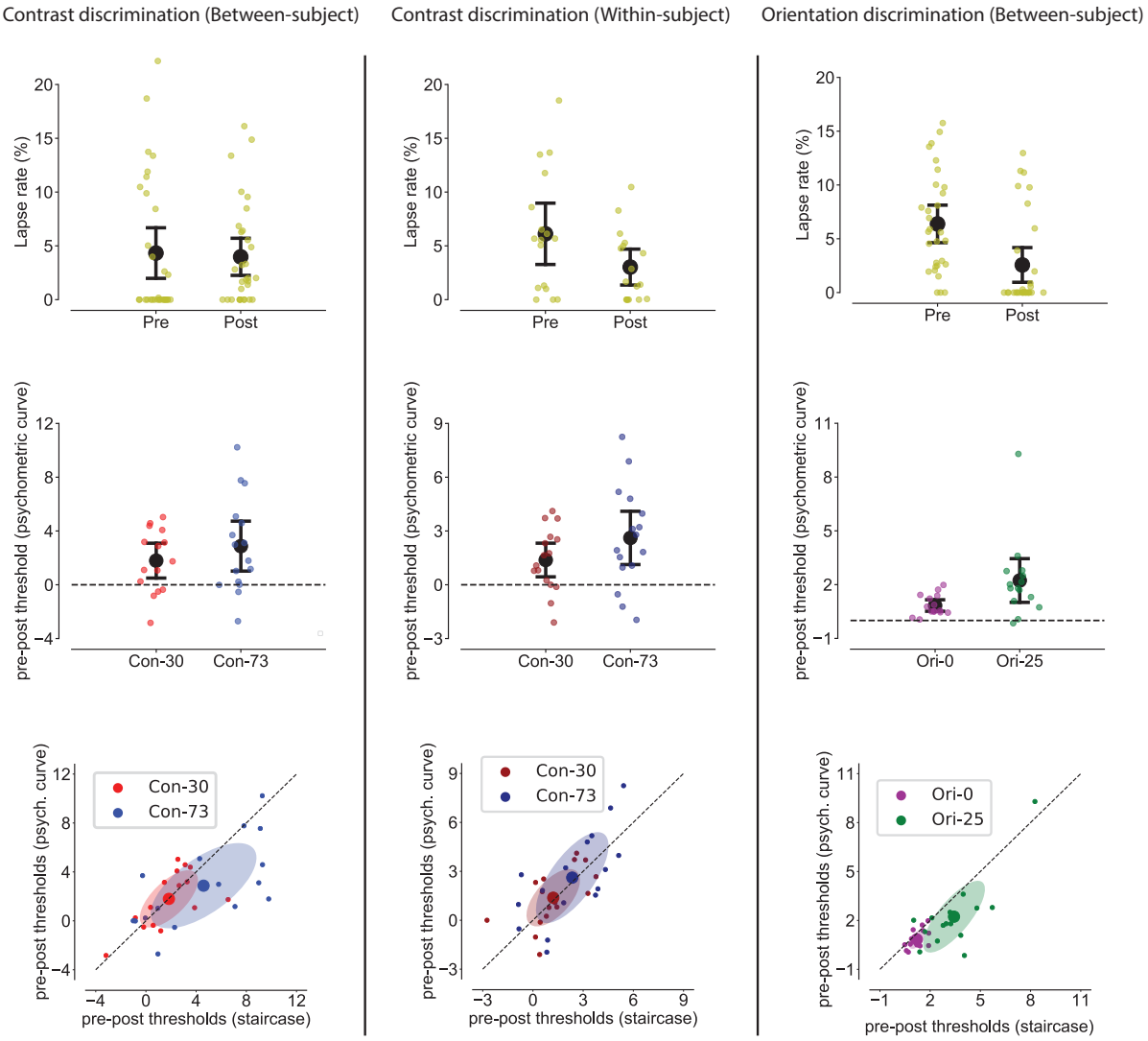


Figure A.1: Top row: The distribution of lapse rates at pre- (before training) and posttests (after training). **Middle row:** The difference between observers' pre- and post-thresholds estimated by the threshold parameter (α) of the best-fitting psychometric curve. **Bottom row:** The improvement in discrimination thresholds due to learning using the reversal points from the adaptive staircase (x axis) is compared to the improvement in discrimination thresholds estimated by the threshold parameter (α) of the best-fitting psychometric curves of the participants at pre- and posttest (y axis). **Left column:** contrast discrimination task, between-subject design. **Middle column:** contrast discrimination task, within-subject design. **Right column:** orientation discrimination task, between-subject design. Error bars represent 95% confidence intervals of the mean. Error ellipses show one standard deviation and the dashed lines mark the x=y values. Adapted with permission from J. Fiser and G. Lengyel.

Although the observers' thresholds decreased significantly due to learning even when we estimated with psychometric curves instead of the reversal points of the staircase method, this will not solve our problem of knowing whether the improvement reflects a decrease in thresholds or a decrease in lapse rates. A reduction in the lapse rate would shift the whole psychometric curve up which in many cases would also result in a decrease in the value of the threshold parameter. To test whether there was any decrease in the thresholds beyond the decrease of the lapse rates, we computed a hypothetical psychometric curve for each subject by adding the amount of decrease in lapse rate due to training to each of the data point of the psychometric curve fitted to the pre-training performance. This method shifted the participants' pre-training psychometric curves up by as much as their lapse-rates decreased after the training and thus, this hypothetical psychometric curve represents approximately the improvement that would have been caused by only improving in lapse rates. We compared the thresholds of the post-training (best-fitting) true psychometric curves to the thresholds of the hypothetical (best-fitting) psychometric curves that assume only lapse rate improvement. We found that thresholds after the training were significantly lower than the corresponding thresholds of the hypothetical psychometric curves that represented the threshold values had they been solely under the control of the decreases in the lapse rates (Fig. A.2). In the orientation discrimination experiment we obtained t_{14}=5.834, P<0.001, d=1.559 with reference orientation 0°, and t_{14}=3.319, P=0.005, d=0.882 with the reference orientation 25° (Fig. A.2, first row, on the right). In the contrast discrimination experiment with between-subject design we obtained t_{14}=3.271, P=0.006, d=0.874, at reference contrast 30%, and t_{15}=3.567, P=0.003, d=0.921 at reference contrast 73% (Fig. A.2, first row, on the left). In the contrast discrimination experiment with within-subject design we obtained t_{15}=2.709, P=0.016, d=0.699, at reference contrast 30%, and t_{15}=3.857, P=0.002, d=0.996 at reference contrast 73% (Fig. A.2, first row, in the middle). Furthermore, this threshold improve-

ment that was controlled for the decrease in lapse rate significantly correlated with the threshold improvement measured by the reversal points from the staircase procedure. These correlations were r=0.58, P<0.001, CI_{95}=0.27-0.78 in the contrast experiment with between-subject design (Fig. A.2, second row, on the left), r=0.68, P<0.001, CI_{95}=0.43-0.84 in the contrast experiment with within-subject design (Fig. A.2, second row, in the middle), and r=0.74, P<0.001, CI_{95}=0.50-0.87 in the orientation experiment with between-subject design (Fig. A.2, second row, on the right).

These results suggest that the decrease in the thresholds after practice was not solely due to the decrease in the lapse rates, and this improvement in perception can be approximated by computing the geometric mean of the reversal points of the adaptive staircase procedure.

A.2 Extended explanation for testing proportionality

The ratio of the observer's initial discrimination thresholds used for scaling the learning scores ($\frac{IT_{Con30}}{IT_{Con73}}$ in Exp. 2, and $\frac{IT_{Ori0}}{IT_{Ori25}}$ in Exp. 3) characterizes the observer's individual perceptual scaling function at the two measured stimulus base-intensities. Therefore, in the first case (Eq. 2.2) the multiplication of the high-reference-value learning scores with participants' initial threshold ratios scaled down participants' learning with the extend of how much larger their initial discrimination thresholds were at the high reference values compared to the low reference values prior to the practice (Fig. 2.4, subpanel A in all panels). This quantity gave us the predicted amount of learning in the untrained low-reference-value condition which can be compared to the measured absolute learning in the other group practicing with that low-reference-value. If the proportionality rule captured by Eq. 2.1 holds, the predicted low-reference-value learning scores should be indistinguishable from the absolute low-reference-value learning scores. Alternatively, if some additional processes influence learning beyond the observers' perceptual scaling,

and the amount of learning will deviate from proportionality rule, the predicted low-reference-value learning scores should be significantly different from the absolute low-reference-value learning scores. In the second case (Eq. 2.3) the division of the low-reference-value learning scores with the participants' initial threshold ratios scaled up participants' learning score with the extend of how much smaller their initial discrimination thresholds were at the low stimulus intensity compared to those at the high intensity (Fig. 2.4, subpanel B in middle and bottom panels). This quantity gave us the predicted amount of learning in the untrained high-reference-value condition which can be compared to the measured absolute learning in the other group practicing with that high-reference-value. The logic of the comparison of the predicted high-reference-value learning scores to the absolute high-reference-value learning scores is the same as in the previous paragraph.

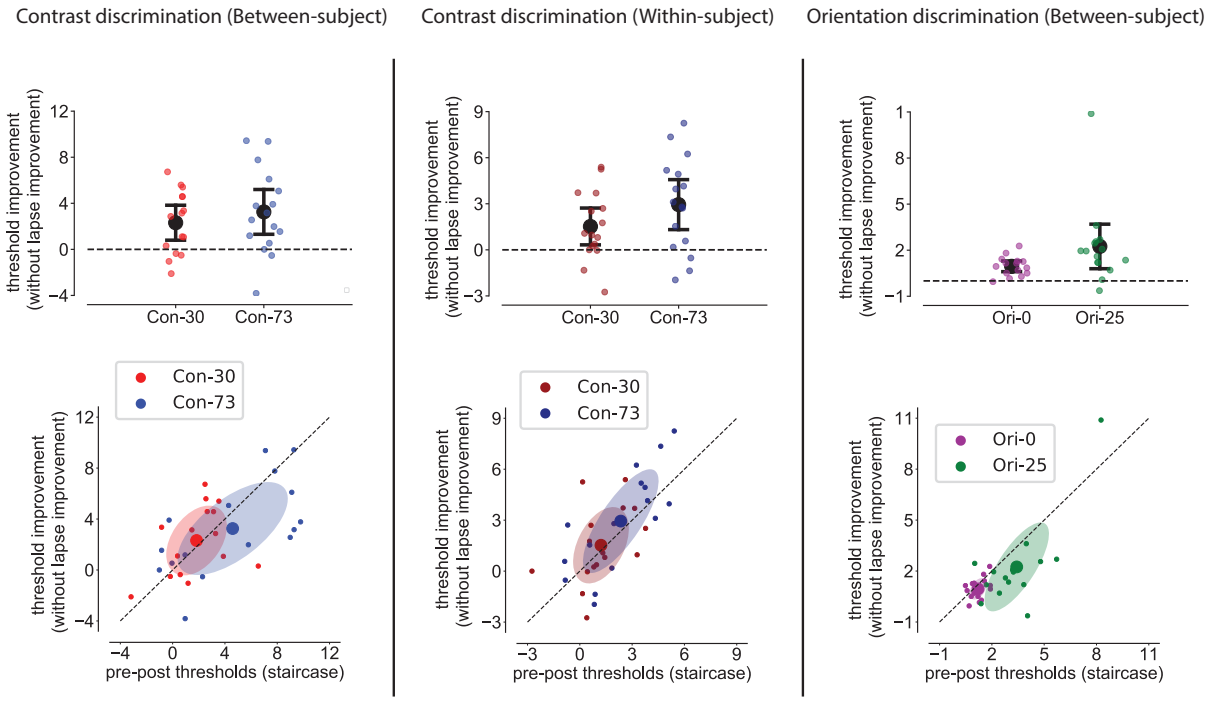


Figure A.2: First row: The decrease in thresholds beyond the decrease in the lapse rates after the training (estimated by best-fitting psychometric curves). **Second row:** The improvement in discrimination thresholds due to learning using the reversal points from the adaptive staircase (x axis) is compared to the improvement in discrimination thresholds beyond the decrease of the lapse rates estimated by best-fitting psychometric curves of the participants' using hypothetical performance assuming only lapse rate decrease due to training and participants' post-training performance (y axis). **Left column:** contrast discrimination task, between-subject design. **Middle column:** contrast discrimination task, within-subject design. **Right column:** orientation discrimination task, between-subject design. Error bars represent 95% confidence intervals of the mean. Error ellipses show one standard deviation and the dashed lines mark the x=y values. Adapted with permission from J. Fiser and G. Lengyel.

A.3 Orientation discrimination experiments

In the orientation discrimination experiments separate groups of observers were trained to discriminate around four different reference values: 0°, 15°, 25°, and 45°. Regarding the investigation of the relationship between initial performance and learning we used 15° and 45° reference values in the first orientation discrimination experiment which did not elicit significant difference in the initial discrimination thresholds. Consequently, we could not test the effect of the different initial performance levels on the amount of learning. In the second experiment we

used 0° and 25° for the orientation references, and we found a large difference between the initial discrimination thresholds, which enabled us to investigate how initial thresholds modulates the amount of learning. In the main text we only reported the results of the latter orientation discrimination experiment. However, in order to show all of our data, we present here all the analysis that we used in the main text for the first orientation discrimination experiment too (in which observers practiced with either 15° or 45° reference values).

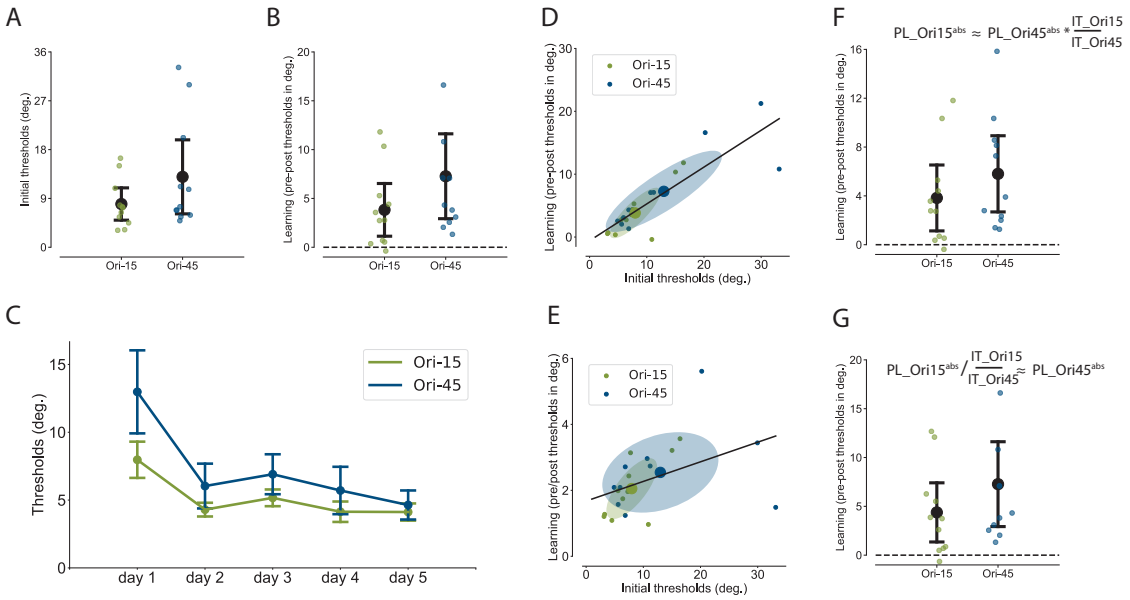


Figure A.3: **(A)** Initial discrimination thresholds and **(B)** the amount of learning at the two measured reference values. **(A, B, F & G)** Error bars represent 95% confidence intervals on the mean. **(C)** Learning curves for the 5-day training protocol for the two measured reference values. Error bars show one SEM. **(D)** Learning as a function of initial discrimination thresholds. **(E)** Relative learning measured as initial discrimination thresholds divided by the post training thresholds as a function of the initial threshold levels. **(D & E)** Error ellipses show one standard deviation, and black lines show linear regression lines fitted to the points from both conditions. **(F)** Comparing the absolute learning in the low reference value condition (gold points) to the predicted learning in the high reference value condition (blue points). **(G)** Comparing the predicted learning in the low reference value condition (gold points) to the absolute learning in the larger reference value condition (blue points). **(F & G)** The equations above the error bars represent the functions of the scaling, where *PL*, absolute learning scores in the specified reference value condition and *IT* denotes the initial thresholds at the specified reference values. Adapted with permission from J. Fiser and G. Lengyel.

Although the initial thresholds seem higher at 45° than at 15° the difference did not reach significance (t_{20}=1.500, P=0.149, d=0.670, Fig. A.3A). There was significant perceptual learn-

ing in both conditions (P<0.05, Fig. A.3B & C) but, the amount of learning did not differ in the two groups (t_{20}=1.499, P=0.150, d=0.670, Fig. A.3B). We computed the predicted learning scores in the group which practiced with 45° reference value using Eq. 2.2 as: $PL^{abs}_{Ori45}\,\frac{IT_{Ori15}}{IT_{Ori45}}$ (see 2.2.3 the main Results of study 1 for more information). When we compared the predicted learning to the absolute learning scores in the group which practiced with 15° reference orientation the difference was not significant (t_{20}=1.067, P=0.299, d=0.477, Fig. A.3F). Similarly, the predicted learning in the group which practiced with the 15° reference values was computed using Eq. 2.3 as: $PL^{abs}_{Ori15}\,/\,\frac{IT_{Ori15}}{IT_{Ori45}}$, and it did not differ significantly from the absolute learning scores in the 45° reference group (t_{20}=1. 217, P=0.238, d=0.544, Fig. A.3G). In terms of the inter-subject variability, there was a large positive correlation between the amount of learning and the initial threshold levels (r=0.85, P<0.001, CI=0.66-0.94, Fig. A.3D). The correlation between relative learning ($PRE/POST$ thresholds) and the initial threshold levels was also significant (r=0.44, P=0.039, CI_{95}=0.12-0.73, Fig. A.3E) but smaller than the correlation between absolute learning and the initial thresholds (z =2.684, p =0.007). These results are in line with the results and the conclusion of the main text.

Regarding the generalization of learning, the inter-subject variability was much smaller with reference orientation 0° than with all other reference orientations (see Fig. 2.3, bottom panels, dots in purple, and Fig. A.5G & H). Therefore, in the second orientation discrimination experiment the correlations between the amount of learning and the extent of generalization gave an unreliable estimate of the true linear relationship between generalization and learning due to the large differences in the variances of the two random variables. Here the two random variables were (1) learning at 0° and generalization at 25°, and (2) learning at 25° and generalization at 0°. Thus, we used the first orientation discrimination experiment with 15° and 45° reference values in the analysis investigating the relationship between learning and generalization (see

2.2.3 Results in the main text for more information).

A.4 Analyzing the amount of learning from the second day

A potential problem weakening the measurement of generalization emerges when no learning took place from Day 2 to Day 5. In this case, learning in the untrained conditions (i.e. improvement at untrained reference values) does not necessarily indicate generalization since the improvement in the trained conditions could be due to the pretest during which observers completed the same amount of trials in the trained and in the untrained conditions. To eliminate this problem, we tested whether there was further improvement in the experiments after the second day of practice and we found that there was significant learning after the second day in most of the conditions. Specifically, we found significant learning from Day 2 to Day 5 in the orientation discrimination experiments (Fig. A.4, bottom panels) at reference orientation 15° (t_{10}=3.05, P=0.01), 45° (t_{10}=3.87, P=0.003), 25° (t_{14}=2.64, P=0.02), and non-significant learning at 0° (t_{14}=0.72, P=0.48). We also found significant, and marginally significant learning from Day 2 to Day 5 in most of the conditions of the contrast discrimination experiments (Fig. A.4, top panels). Specifically, we obtained t_{23}=1.74, P=0.09 and t_{22}=2.57, P=0.01 in the contrast experiment with between-subject design at reference con. 30% and 73% respectively. In the contrast experiment with within-subject design we found t_{16}=0.37, P=0.71 and t_{16}=1.77, P=0.09 at reference con. 30% and 73% respectively.

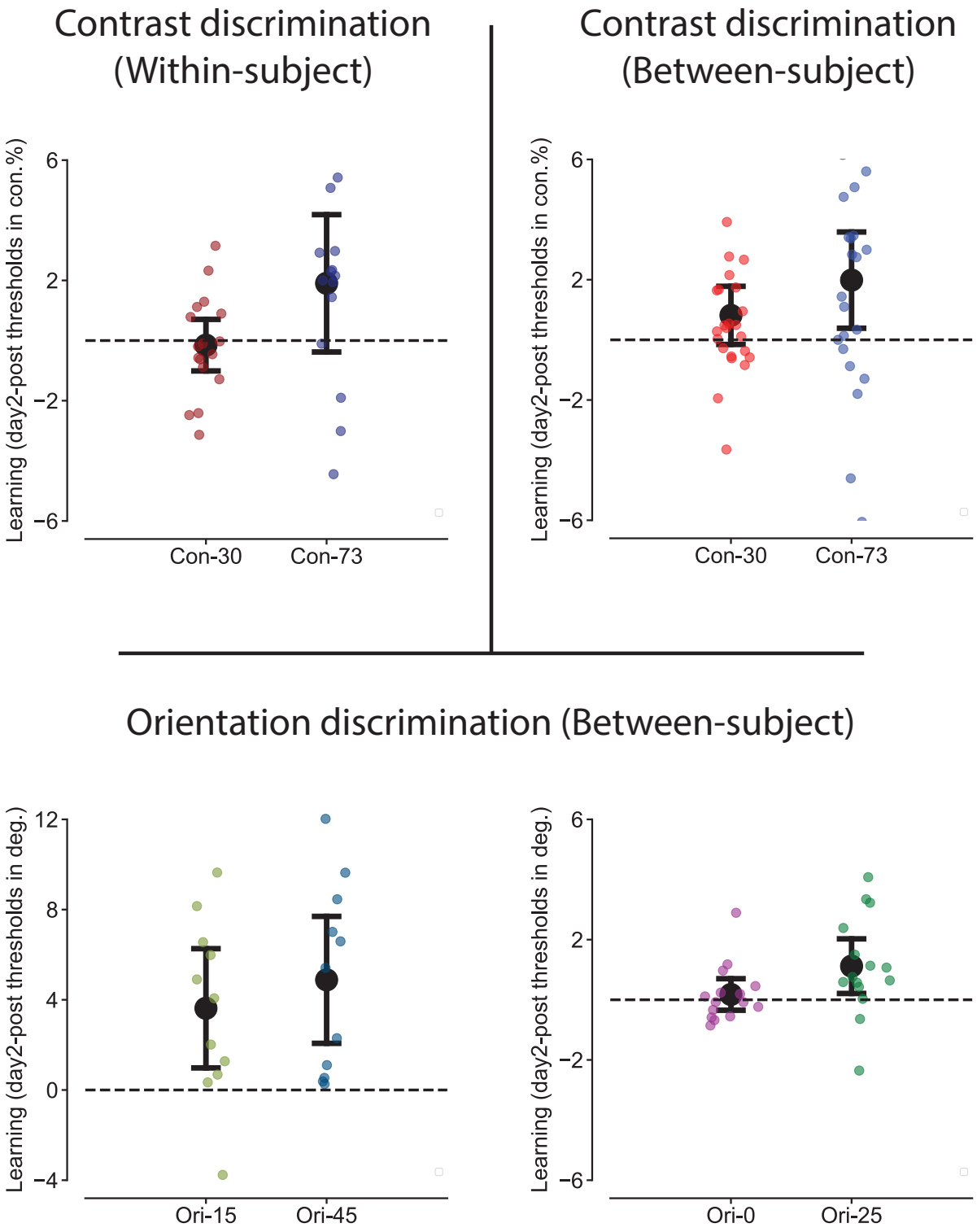


Figure A.4: The amount of learning between Day 2 and the final posttest. **Top panel:** contrast discrimination task, within-subject design. **Middle panel:** contrast discrimination task, between-subject design. **Bottom panel:** orientation discrimination task, between-subject design. Error bars represent 95% confidence intervals of the mean. Adapted with permission from J. Fiser and G. Lengyel.

There was no improvement after the second day in the cardinal (0°) reference orientation condition in the orientation discrimination task (Fig. A.4, bottom right). However, we did not use this group of observers in the analysis for generalization of learning anyway because of its excessively small inter-subject variability (see 2.2.3 results, learning and generalization and Appendix A A.3, orientation discrimination experiments for more detail). Regarding the contrast experiment with within-subject conditions (in which observers show the lowest amount of improvement from Day 2 across experiments) we cannot conclude that there was no further improvement from Day 2 in this condition either. This is because there was a marginally significant improvement in the condition with reference contrast 73%, and only 3 subjects showed no improvement from Day 2 (Fig. A.4, top left). Thus, learning could have transferred from that condition to the untrained middle reference value con. 47% in most participants.

A.5 Extended analysis of learning and generalization

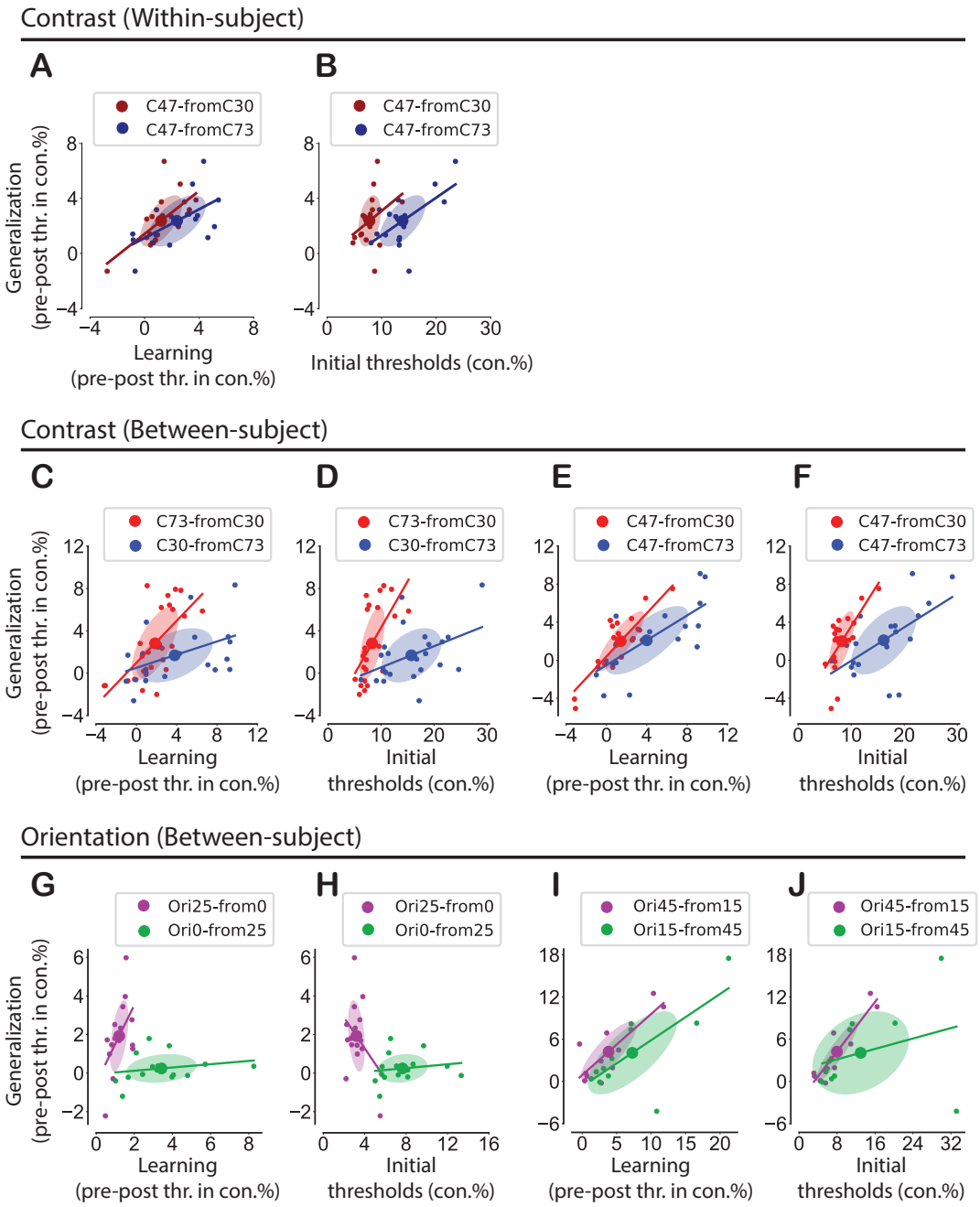


Figure A.5: Top panel: contrast discrimination task, within-subject design. **Middle panel:** contrast discrimination task, between-subject design. **Bottom panel:** orientation discrimination task, between-subject design. **(A, C, E, G & I):** Generalization as a function of learning. **(B, D, F, H & J):** Generalization as a function of initial discrimination thresholds. In all plots error ellipses show one standard deviation and colored lines represent linear regression lines for the corresponding conditions. The first part of the labels (C73-, C30-, C47-, Ori0-, Ori25-) denotes the reference value at which the generalization was measured, while the second part of the labels (-fromC73, -fromC30, -from25, -from0) denotes the practiced reference values from which the learning transferred. Adapted with permission from J. Fiser and G. Lengyel.

Correlations between learning and generalization			
Experiment	Correlation coefficient	95% Confidence interval	p value
Exp 1. transfer to con. 47% from con. 30% (Fig. S4 A, red line)	r = 0.64	CI95 = 0.23 - 0.87	p = 0.005
Exp 1. transfer to con. 47% from con. 73% (Fig. S4 A, blue line)	r = 0.58	CI95 = 0.13 - 0.84	p = 0.013
Exp 2. transfer to con. 73% from con. 30% (Fig. S4 C, red line)	r = 0.66	CI95 = 0.34 - 0.85	p < 0.001
Exp 2. transfer to con. 30% from con. 73% (Fig. S4 C, blue line)	r = 0.46	CI95 = 0.05 - 0.74	p = 0.027
Exp 2. transfer to con. 47% from con. 30% (Fig. S4 E, red line)	r = 0.85	CI95 = 0.63 - 0.94	p < 0.001
Exp 2. transfer to con. 47% from con. 73% (Fig. S4 E, blue line)	r = 0.74	CI95 = 0.43 - 0.90	p < 0.001
Exp 3. transfer to ori. 25˚ from ori.0˚ (Fig. S4 G, purple line)	r = 0.53	CI95 = 0.00 - 0.82	p = 0.044
Exp 3. transfer to ori.0˚ from ori.25˚ (Fig. S4 G, green line)	r = 0.21	CI95 = -0.35 - 0.66	p = 0.450
Exp 3. transfer to ori. 15˚ from ori.45˚ (Fig. S4 I, green line)	r = 0.82	CI95 = 0.40 - 0.95	p = 0.002
Exp 3. transfer to ori. 45˚ from ori.15˚ (Fig. S4 I, purple line)	r = 0.72	CI95 = 0.19 - 0.92	p = 0.013
Correlations between initial thresholds and generalization			
Exp 1. transfer to con. 47% from con. 30% (Fig. S4 B, red line)	r = 0.36	CI95 = -0.16 - 0.73	p = 0.154
Exp 1. transfer to con. 47% from con. 73% (Fig. S4 B, blue line)	r = 0.62	CI95 = 0.18 - 0.85	p = 0.008
Exp 2. transfer to con. 73% from con. 30% (Fig. S4 D, red line)	r = 0.61	CI95 = 0.26 - 0.82	p = 0.002
Exp 2. transfer to con. 30% from con. 73% (Fig. S4 D, blue line)	r = 0.44	CI95 = 0.02 - 0.73	p = 0.034
Exp 2. transfer to con. 47% from con. 30% (Fig. S4 F, red line)	r = 0.67	CI95 = 0.30 - 0.87	p = 0.002
Exp 2. transfer to con. 47% from con. 73% (Fig. S4 F, blue line)	r = 0.60	CI95 = 0.18 - 0.82	p = 0.008
Exp 3. transfer to ori. 25˚ from ori.0˚ (Fig. S4 H, purple line)	r = -0.35	CI95 = -0.74 - 0.21	p = 0.20
Exp 3. transfer to ori.0˚ from ori.25˚ (Fig. S4 H, green line)	r = 0.16	CI95 = -0.40 - 0.63	p = 0.565
Exp 3. transfer to ori. 15˚ from ori.45˚ (Fig. S4 J, green line)	r = 0.91	CI95 = 0.68 - 0.98	p = 0.002
Exp 3. transfer to ori. 45˚ from ori.15˚ (Fig. S4 J, purple line)	r = 0.32	CI95 = -0.37 - 0.78	p = 0.341

Figure A.6: Analyzing the linear relationship between the extent of generalization, the amount of learning, and the initial discrimination thresholds. Adapted with permission from J. Fiser and G. Lengyel.

Appendix B

Supplementary materials for study 2

B.1 Experiment 1

B.1.1 Descriptive statistics

The median reaction time and mean error rates of the observers for the three types of response in all blocks are shown in Fig. B.1 for the main (a), replication (b), and in the control (c) experiments. Although the task was difficult and produced relatively high error rates, observers paid attention to the task as indicated by the longer search times in each experiment in trials with only one target letter T. In the following, R1 refers to trials with two target letter Ts appearing vertically arranged on top of each other, R2 refers to trials with two target letter Ts appearing horizontally arranged next to each other, and R3 denotes trials with only one target appearing.

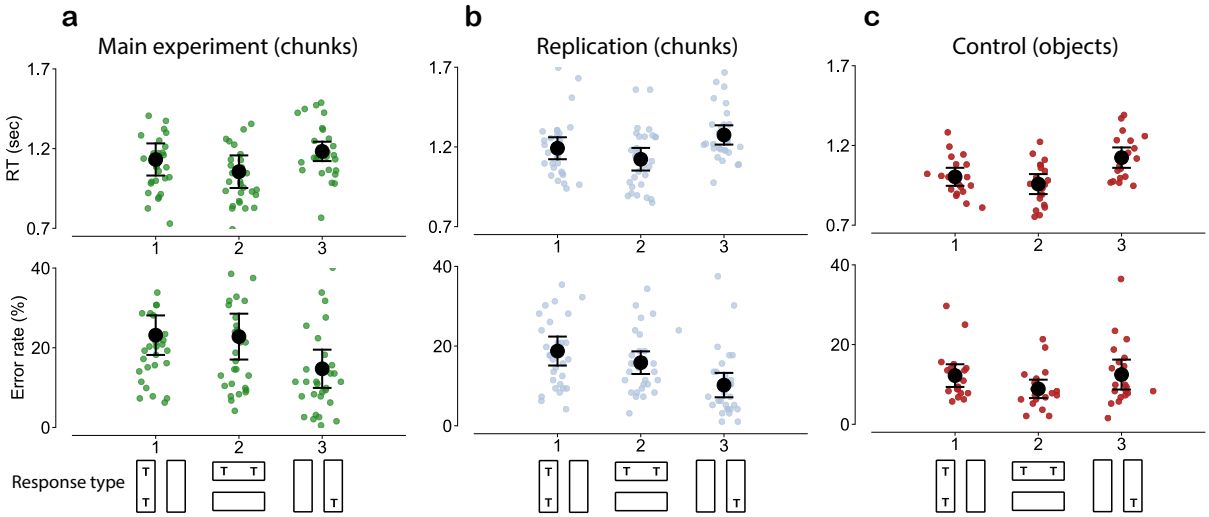


Figure B.1: Median reaction times (top) and mean error rates (bottom) for the three response types in the main (Experiment 1a, column a), in the replication (Experiment 1b, column b), and in the control tests (Experiment 1c, column c). The response types are shown on the x axis: (1) two target letter Ts appearing vertically arranged on top of each other, (2) two target letter Ts appearing horizontally arranged next to each other, and (3) only one target appearing. Error bars in all plots show 95% confidence intervals of the mean. Colored dots represent the mean error rates or median reaction time of the observers. n=30 in Exp. 1a (a), n=30 in Exp. 1b (b), and n=20 in Exp. 1c (c). Adapted with permission from G. Lengyel, M. Nagy, and J. Fiser.

Reaction times of the observers to R1, R2 and R3 trials differed significantly in all three experiments. In the main experiment: $F_{2,58}$=6.195, P<0.004, Bayes Factor=10, post-hoc comparing R1 and R2, t_{29}=3.82, P<0.001, d=0.709, Bayes Factor=49, R1 and R3, t_{29}=1.24, P=0.226, d=0.230, Bayes Factor=0.4, and R3 and R2, t_{29}=2.92, P=0.006, d=0.542, Bayes Factor=6 (Fig. B.1a, top). In the replication experiment: $F_{2,58}$=21.42, P<0.001, Bayes Factor=105, post-hoc comparing R1 and R2, t_{29}=3.59, P=0.001, d=0.667, Bayes Factor=28, R1 and R3, t_{29}=2.98, P=0.006, d=0.554, Bayes Factor=7, and R3 and R2, t_{29}=6.95, P<0.001, d=1.290, Bayes Factor=$1.2 \cdot 10^5$ (Fig. B.1b, top). In the control experiment: $F_{2,38}$=33.6, P<0.001, Bayes Factor=$1.9 \cdot 10^6$, post-hoc comparing R1 and R2, t_{19}=2.36, P=0.029, d=0.540, Bayes Factor=2, R1 and R3, t_{19}=6.05, P<0.001, d=1.390, Bayes Factor=2685, and R3 and R2, t_{19}=7.13, P<0.001, d=1.640, Bayes Factor=$2 \cdot 10^4$ (Fig. B.1c, top).

Observers' error rates to R1, R2 and R3 trials were also significantly different in all three

experiments. In the main experiment: $F_{2,58}$=9.26, P<0.001, Bayes Factor=83, post-hoc comparing R1 and R2, t_{29}=0.197, P=0.845, d=0.037, Bayes Factor=0.2, R1 and R3, t_{29}=3.95, P<0.001, d=0.733, Bayes Factor=66, and R3 and R2, t_{29}=3.11, P<0.001, d=0.577, Bayes Factor=9 (Fig. B.1a, bottom). In the replication experiment: $F_{2,58}$=13.04, P<0.001, Bayes Factor=1184, post-hoc comparing R1 and R2, t_{29}=2.09, P=0.046, d=0.388, Bayes Factor=1, R1 and R3, t_{29}=4.35, P<0.001, d=0.808, Bayes Factor=176, and R3 and R2, t_{29}=3.32, P=0.002, d=0.617, Bayes Factor=15 (Fig. B.1b, bottom). In the control experiment: $F_{2,38}$=3.18, P=0.053, Bayes Factor=1, post-hoc comparing R1 and R2, t_{19}=3.52, P=0.002, d=0.808, Bayes Factor=18, R1 and R3, t_{19}=0.14, P=0.890, d=0.0321, Bayes Factor=0.2, and R3 and R2, t_{19}=1.99, P=0.060, d=0.459, Bayes Factor=1, Fig. B.1c, bottom).

These statistical analyses confirm two expected outcomes. First, observers were faster when the two targets appeared next to each other as opposed to when the targets were on top of each other. This could reflect a bias effect to the horizontal reading direction. Second, observers searched longer when there was only one target letter confirming the effect of observers' extended search for a second target. Observers also made fewer error when there was only one target in the main and the replication experiment but not in the control experiment. This is likely due to the lower base error rate in the control experiment.

B.1.2 Diminishing chunk- and object-based effects

More importantly, we also hypothesized that the object- and chunk-based effects in error rates would be large initially, when observers commit many errors in the process of learning the task, and the effect would decrease significantly later, when they reach a good performance with fewer errors. To evaluate this hypothesis, we tested whether observers' error rates dropped after the first block, that is whether their performance increased significantly. Indeed, we found

that observers made more errors and responded slower in the first block compared to the other blocks. One-way ANOVA of error rates in Experiment 1a showed a main effect of blocks ($F_{3,203}$=31.58, P<0.001, Bayes Factor=$4.5 \cdot 10^{1}$3), and post-hoc comparisons of block 1 to block 2-4 confirmed a significant difference (ts_{29}>3.38, Ps<0.002, Bayes Factors>17) (Supplementary Fig. 2a). The same analysis for reaction times also found a main effect of blocks ($F_{3,203}$=94.438, P<0.001, Bayes Factor=$1.5 \cdot 10^{3}$5), and a significant post-hoc differences when comparing block 1 to block 2-4 (ts_{29}>6.88, Ps<0.001, Bayes Factors>$1.1 \cdot 10^{5}$) (Fig. B.2d).

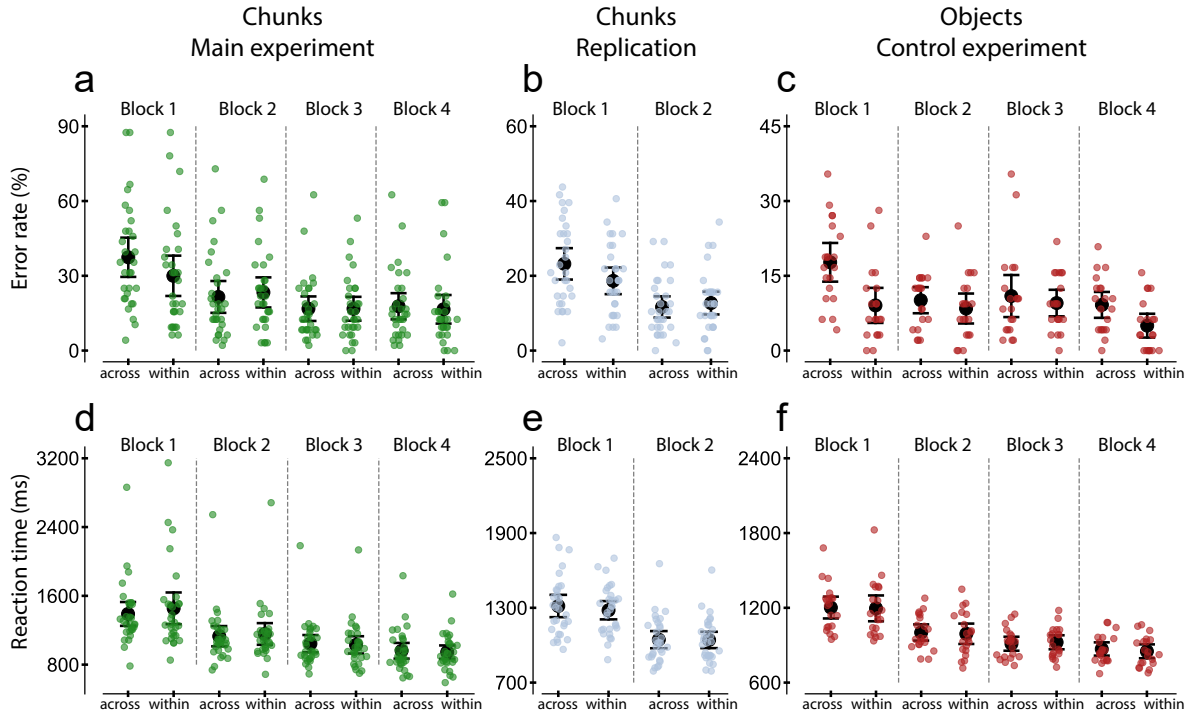


Figure B.2: Mean error rates (a-c) and median reaction times (d-f) in Experiment 1a (a, d), Experiment 1b (b, e), and in Experiment 1c (c, f) for each block (separated by dotted lines) in the across- and within-chunk/object conditions. On the x axis, the across condition denotes trials, in which the two target letters Ts appeared across two chunks/objects, while the within condition represents trials, in which the two targets Ts were confined to a single chunk/object. Dots represent individual observers' performance; error bars indicate the 95% confidence intervals of the mean. Note the different scales on the y axes. n=30 in Exp. 1a (a, d), n=30 in Exp. 1b (b, e), and n=20 in Exp. 1c (c, f). Adapted with permission from G. Lengyel, M. Nagy, and J. Fiser.

The analysis of Experiment 1b yielded the same results. Blocks had a main effect on both error rates ($F_{3,203}$=66.57, P<0.001, Bayes Factor=$\cdot 10^{8}$), with significant advantage of the first block (post-hoc comparing block 1 to block 2, t_{29}=7.79, P<0.001, Bayes Factor=$\cdot 105$)

(Fig. B.2b), and on reaction times ($F_{3,203}$=184.731, P<0.001, Bayes Factor=$1.1 \cdot 10^20$), with a significant disadvantage of block 1 over block 2 (t_{29}=10.39, P<0.001, Bayes Factor=$3.4 \cdot 10^8$) (Fig. B.2e).

The control experiment (1c) yielded the same pattern of results as the chunk-based experiments. There was a main effect of blocks both in error rates, ($F_{3,133}$=11.77, P<0.001, Bayes Factor=1543), dominated by a significantly larger error in block 1 compared to block 2-4 (ts_{19}>2.09, Ps<0.05, Bayes Factors>1), (Fig. B.2c), and in reaction times ($F_{3,133}$=71.56, P<0.001, Bayes Factor=$1.1 \cdot 10^25$), again mostly due to the significantly slower RTs in block 1 comparing to block 2-4 (ts_{19}>6.24,Ps<0.001, Bayes Factors>3817) (Fig. B.2f).

These results clearly support the hypothesis that in every experiment and both in error rates and reaction times, the largest improvement took place between the first and the second block and in subsequent blocks the improvement was negligible. This floor effect diminishing the potential difference between the within- and between-unit effects explains why the chunk/object-based effect in all experiments have disappeared after the first block.

B.2 Experiment 2

B.2.1 Constructing the catch trials in the familiarity test

For each observer, two foil pairs, one horizontal and one vertical, were generated randomly from the diagonal pairs the same way as in Experiment 1. We refer to these two foil pairs as diagonal-pair foils. More foil pairs were created from the true-pairs in the following way. Two horizontal-pair foils were created by pairing the two top and the two bottom shapes of the two vertical true pairs. Two vertical-pair foils were created by pairing the two left and the two right shapes of the two horizontal true-pairs. We call these four foil pairs true-pair foils.

Testing the true-pair foils against the true-pairs contrasts directly the true-pairs against those shape combinations that occurred when two true-pairs were put together side by side in the search trials creating possible pairs orthogonal to the boundary of the true-pairs.

Finally, two additional foil pairs were created from the diagonal-pair foils in the following way. The top shape of the vertical diagonal-pair foil was paired with the left shape of the horizontal diagonal-pair foil forming a vertical foil pair. The last, horizontal foil pair was created by pairing the bottom shape of the vertical with the right shape of the horizontal diagonal-pair foils. In the 8 catch trials the 4 true-pair foils were tested against these two additional foil pairs.

B.2.2 Average RTs and errors with chunks and objects

We compared the reaction times (RTs) and error rates (ERs) of the object and chunk version of the paradigm (Fig. B.3). Observers were faster (t_{43}=6.02, P<0.001, d=0.918, Bayes Factor=$4.6 \cdot 10^4$) and made fewer error (t_{43}=3.84, P<0.001, d=0.585, Bayes Factor=67) in the paradigm using objects (rectangles). Some part (or all) of this effect could be explained by a generic, maybe attentional based, learning effect because the object version of the paradigm always appeared after all other tasks (the 4-4 blocks of VSL and CBA and the familiarity test). Nevertheless, observers' behavior was very similar across the two stimulus sets.

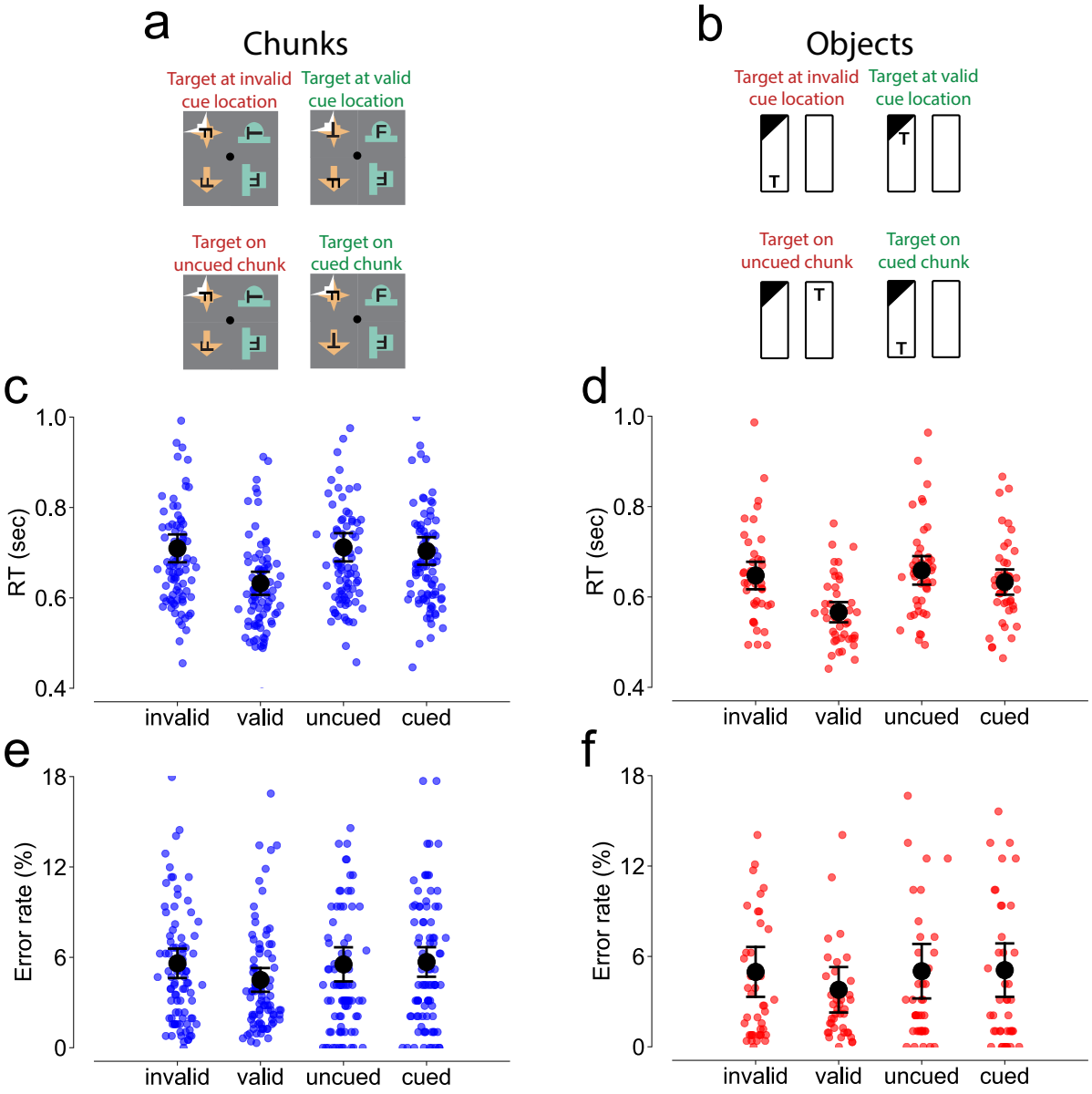


Figure B.3: Median reaction times (c, d) and mean error rates (e, f) in the four conditions of Experiment 2 using chunks (a, c, e) and objects (b, d, f). Labels on the x axes: invalid - the target appeared at an uncued location; valid - the target appeared at the cued location; uncued - within the invalid-cue trials, the target appeared on the uncued chunk; cued - within the invalid-cue trials, the target appeared on the cued chunk. Error bars indicate the 95% confidence intervals of the mean; dots represent individual observers' performance. n=90 in the blocks with statistical chunks (a, c, e in blue), and n=44 in the blocks with geometric objects (b, d, f in red). Adapted with permission from G. Lengyel, M. Nagy, and J. Fiser.

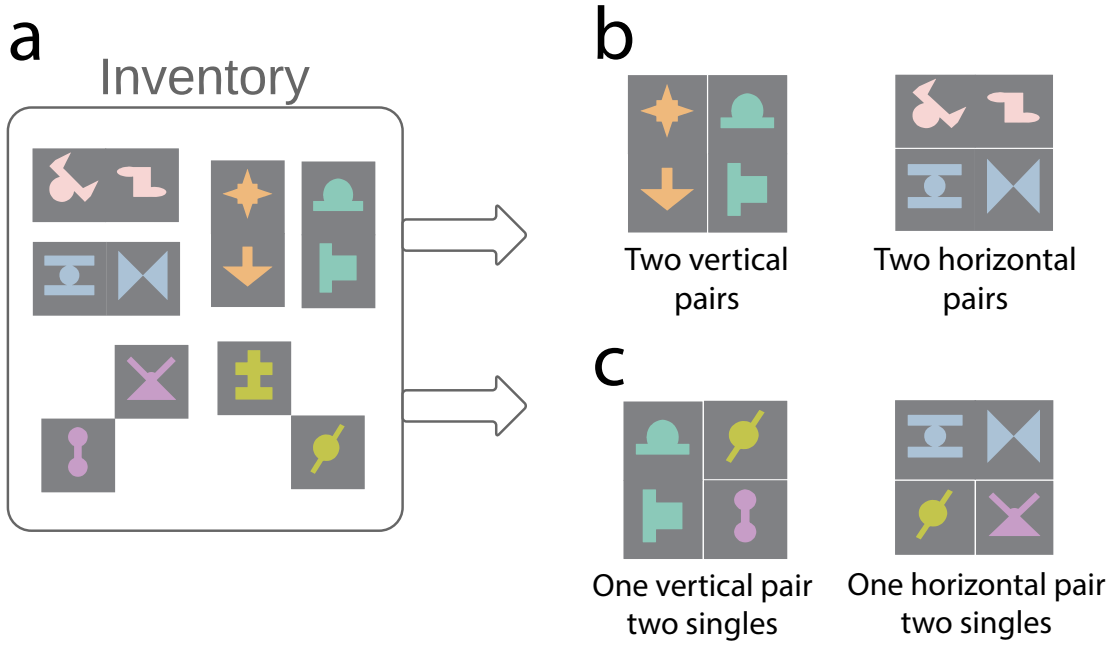


Figure B.4: Shows an example of an inventory from which the true-pairs were generated throughout the experiments. b, c Show the four types of trials in the search tasks in all experiments. In this figure, colors are just for demonstration and in the experiments all shapes were black, smaller that in this figure, and were separated by back lines in experiment 1a & b (see Fig. 1b). The separating back lines were removed in experiment 2 (see Fig. 3a). In half of the search trials two true-pairs, either vertically or horizontally oriented, were presented (see two example scenes in b). In the other half of the search trials one true-pair, either vertically or horizontally oriented, and two individual shapes from the cross-pairs were presented (see two example scenes in c). Note that we do not show here all the unique 2-by-2 scenes for the four trial types that one could generate from the Inventory. However, it is easy to see that only 4 unique 2-by-2 scenes can be generated from the two vertically and two horizontally oriented true-pairs, and 96 unique scenes can be generated from one vertically or horizontally oriented true-pairs and from two individual shapes from the cross-pairs. We presented all 96 unique scenes containing one true-pair and two individual shapes and 12 times the 4 unique scenes containing only two true-pairs. Since in half of the trials containing individual shapes the targets appeared on the individual shape we had the same amount of trials containing two true-pairs and one true-pair and two individual shapes for measuring object-based effects. Adapted with permission from G. Lengyel, M. Nagy, and J. Fiser.

Appendix C

Supplementary materials for Chapter 4

C.1 The encoding-decoding framework

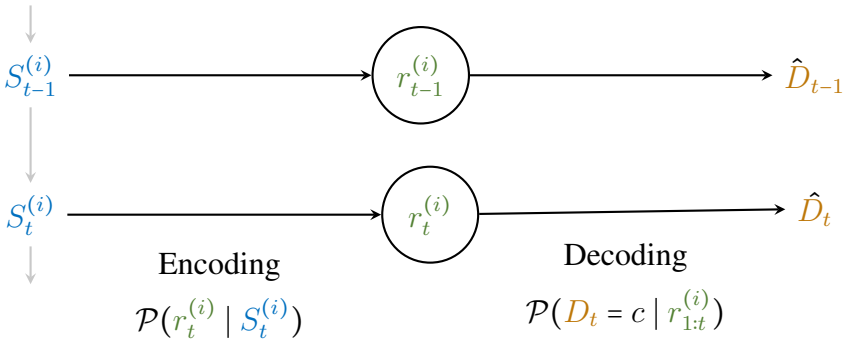


Figure C.1: Encoding-decoding framework. Perception is captured as an encoding-decoding process. $S_t^{(i)}$ represents the stimulus, $r_t^{(i)}$ denotes the population response of the sensory neurons to the stimulus, $\hat{D}_t$ represents the decoded decision category in trial t, and $i \in \{1, \ldots, K\}$ where K is the number of different stimuli or stimulus features in a trial.

C.2 The Fisher information

Let's assume that the population responses of the sensory neurons to the stimulus, denoted by r, can be formalized as independent Poisson random variables:

$$r \mid S \sim \prod_{n=1}^{N} \text{Poisson}\{f_n(S)\} \tag{C.1}$$

where N denotes the number of neurons in the population and the expected activity of the nth neuron in response to stimulus $S = s$ is represented by the tuning function of the nth neuron, $f_n(S = s)$. Most computational models in PL assume that there are sensory units/neurons with hypothetical tuning functions, describing the mean activity rates, and with multiplicative Gaussian (e.g., Talluri et al., 2015) or Poisson (e.g., Jazayeri and Movshon, 2006) noise. The function f is usually formalized based on the average firing rates of experimentally measured sensory neurons to a stimulus set.

From the observer point of view, r is the sensory observations generated by the stimuli, S. In case of the independent Poisson random variables given in Eq. C.1 and in the limit of larger N the generative model of the sensory observations can be written as abstract, Gaussian random variables using the Fisher information (Seung & Sompolinsky, 1993):

$$\tilde{S} \mid S \sim \mathcal{N}\left(S, \frac{1}{I_{\text{F}}(S)}\right) \tag{C.2}$$

where $\tilde{S}$ represents a single, abstract, sensory observation given the stimuli, and $I_{\text{F}}(S)$ denotes the Fisher information which quantifies the amount of information in the population response, r, about the stimulus value S, defined as follows (Cox & Hinkley, 1974):

$$I_{\text{F}}(S) = -\int \mathcal{P}(r \mid S) \frac{\partial^2}{\partial S^2} \ln \mathcal{P}(r \mid S) \mathrm{d}\, r \tag{C.3}$$

In case of independent Poisson distributions the Fisher information can be computed from the tuning curves of the neurons (Seung & Sompolinsky, 1993):

$$I_{\mathrm{F}}(S) = \sum_{n=1}^{N} \frac{f_n'^2(S)}{f_n(S)} \tag{C.4}$$

where $f_n'^2(S)$ denotes the derivative of the nth tuning curve. All of the characteristics of the population response in Eq. C.1, in the limit of larger N, is captured by the Fisher information in Eq. C.3.

C.3 Short description of Ganguli and Simoncelli (2014)

Ganguli and Simoncelli (2014) derived a closed-form solution for optimally allocating sensory neurons (parameterized with a density function) and spikes (parameterized with a gain function) to maximize discrimination performance given a prior distribution over the stimulus values, a resource constraint, and a reasonable approximation of the Fisher information. Using their closed-form solution, I can compute directly the optimal tuning properties of the sensory neurons (i.e., the optimal encoding) given the probability of the stimuli in a PL experiment to maximize performance.

I begin with briefly describing the model that Ganguli and Simoncelli (2014) developed. Similar to other studies (Jazayeri & Movshon, 2006; Seriès et al., 2009; Seung & Sompolinsky, 1993), the authors assumed that there is a single, homogeneous population of N sensory neurons with unimodal or sigmoidal response profiles tuned to different feature values spanning the whole feature space uniformly. The number of spikes the neurons emit in a given period of time is generated from independent Poisson distributions with mean activities described by the tuning curves of the neurons (see Eq. C.1). Furthermore, the total expected spike rate of the

neuron population, denoted by R, is assumed to be fixed constraining the tuning curves:

$$\int \mathcal{P}(S) \sum_{n=1}^{N} f_n(S) \mathrm{d}S = R \tag{C.5}$$

where $\mathcal{P}(S)$ denotes the probability of the feature values of the stimuli in the experiment. To represent values of the stimulus drawn from the distribution, $\mathcal{P}(S)$, in an efficient way, using N neurons and limiting the total expected spike rate of the population, the tuning curves of the neurons should be adjusted to maximize the mutual information between the stimuli and the population responses. Since maximizing mutual information is computationally expensive, the authors chose to optimize the lower bound on mutual information which can be expressed using the Fisher information (see Brunel and Nadal, 1998):

$$\underset{\{f_n(S)\}}{\arg\max} \int \mathcal{P}(S) \log\left(I_{\mathrm{F}}(S)\right) \mathrm{d}S, \quad \text{s.t.} \quad \int \mathcal{P}(S) \sum_{n=1}^{N} f_n(S) \, \mathrm{d}S = R \tag{C.6}$$

where $I_{\mathrm{F}}(S)$ represent the Fisher information (see Eqs. C.3 and C.4).

The Fisher information can also be used to provide a lower bound on discriminability (Seriès et al., 2009):

$$\delta(S) = \frac{C}{\sqrt{I_{\mathrm{F}}(S)}} \tag{C.7}$$

C represent a constant that is set based on the threshold levels measured in the discrimination experiments. Thus, in order to maximize performance in discrimination the function $\log\{I_{\mathrm{F}}(S)\}$ can be replaced by the $-\left\{\frac{1}{I_{\mathrm{F}}(S)}\right\}$ function in the optimization in C.6 which then will maximize the squared discriminability.

The authors developed a parametric model of the tuning curves in which they formalized the population of tuning curves as a warped and rescaled version of the initial population with identical tuning curves that spans the entire feature space uniformly. The warping and the

scaling transformations are characterized by a density $d(S)$ and a gain $g(S)$ functions. The widths of the tuning curves in the warped and scaled population are proportional to the distances between the tuning curves; the higher the density of the neurons around a feature value the narrower their tuning curves are maintaining the amount of overlaps between the tuning curves the same as in the initial homogeneous tuning curve population. The function, $d(S)$, determines the local allocation of the neurons, while the function, $g(S)$, is responsible for the local allocation of the number of spikes emitted by those neurons (see (Ganguli & Simoncelli, 2014) for specifying the parameterization and for the detailed derivations).

By assuming that the Fisher information is approximately constant in the initial, homogeneous, and uniformly tiled tuning curve population to all feature values and that $g(S)$ is smooth relative to the width of the Fisher information for the single warped neurons (see Ganguli and Simoncelli, 2014 for more details) the Fisher information in the optimization in C.6 can be approximated with the following term: $d^2(S)g(S)$. Using this approximation the optimization for maximizing discrimination becomes:

$$\underset{d(S),g(S)}{\arg\max} \int \mathcal{P}(S) - \frac{1}{(d^2(S)g(S))}\,\mathrm{d}S, \quad \text{s.t.} \tag{C.8}$$

$$\int d(S)\,\mathrm{d}S = N, \quad \text{and} \quad \int \mathcal{P}(S)g(S)\,\mathrm{d}S = R$$

Solving this optimization yields the following optimal solution for unimodal tuning curves (Ganguli & Simoncelli, 2014):

$$d(S) \propto N\sqrt{\mathcal{P}(S)}, \quad g(S) \propto R\frac{1}{\sqrt{\mathcal{P}(S)}} \tag{C.9}$$

Regarding the density function, the solution above is in line with the intuition that allocating

more neurons with narrower tuning curves for feature values that are more probable in the discrimination task results in lower discrimination thresholds. In case of the gain function the optimal solution is to represent feature values with higher probability with lower firing rates. The Fisher information given the optimal solutions can also be approximated using only the probability of the stimulus values:

$$I_F(S) \propto RN^2\sqrt{\mathcal{P}(S)} \tag{C.10}$$

The authors also derived the solution for the optimization in C.8 assuming sigmoidal tuning curves (Ganguli & Simoncelli, 2014):

$$\begin{aligned} d(S) &\propto N\sqrt[3]{\mathcal{P}(S)}\sqrt[3]{1-\int_{-\infty}^{S}\mathcal{P}(S)\,\mathrm{d}S}, \\ g(S) &\propto R\frac{1}{N}\frac{1}{1-\int_{-\infty}^{S}\mathcal{P}(S)\,\mathrm{d}S} \\ I_F(S) &\propto RN\mathcal{P}_t^{\frac{2}{3}}(S)(1-\int_{-\infty}^{S}\mathcal{P}(S)\,\mathrm{d}S)^{-\frac{1}{3}} \end{aligned} \tag{C.11}$$

The equations in C.9 - C.11 show how to allocate the resources in terms of the location of the tuning preferences and number of spikes given the probability of the stimulus feature values.

Bibliography

Abbott, L. & Regehr, W. (2004). Synaptic computation. *Nature*, *431*, 796–803.

Aberg, K. C. & Herzog, M. H. (2009). Interleaving bisection stimuli – randomly or in sequence – does not disrupt perceptual learning, it just makes it more difficult. *Vision Res.*, *49*(21), 2591–2598.

Aberg, K. C. & Herzog, M. H. (2012). Different types of feedback change decision criterion and sensitivity differently in perceptual learning. *J. Vis.*, *12*(3).

Abla, D., Katahira, K. & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *J. Cogn. Neurosci.*, *20*, 952–964.

Adab, H. Z., Popivanov, I. D., Vanduffel, W. & Vogels, R. (2014). Perceptual learning of simple stimuli modifies stimulus representations in posterior inferior temporal cortex. *J. Cogn. Neurosci.*, *26*(10), 2187–2200.

Adab, H. Z. & Vogels, R. (2011). Practicing coarse orientation discrimination improves orientation signals in macaque cortical area v4. *Curr. Biol.*, *21*(19), 1661–1666.

Adini, Y., Sagi, D. & Tsodyks, M. (2002). Context-enabled learning in the human visual system. *Nature*, *415*(6873), 790–793.

Adini, Y., Wilkonsky, A., Haspel, R., Tsodyks, M. & Sagi, D. (2004). Perceptual learning in contrast discrimination: The effect of contrast uncertainty. *J. Vis.*, *4*(12), 993–1005.

Ahissar, M. & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401–406.

Ahissar, M. & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.*, *8*(10), 457–464.

Aitchison, L., Jegminat, J., Menendez, J. A., Pfister, J.-P., Pouget, A. & Latham, P. E. (2021). Synaptic plasticity as bayesian inference. *Nature Neuroscience*, *24*(4), 565–571. https://doi.org/10.1038/s41593-021-00809-5

Aitchison, L. & Lengyel, M. (2017). With or without you: Predictive coding and bayesian inference in the brain. *Curr. Opin. Neurobiol.*, *46*, 219–227.

Alamia, A., Solopchuk, O., D'Ausilio, A., Van Bever, V., Fadiga, L., Olivier, E. & Zénon, A. (2016). Disruption of broca's area alters higher-order chunking processing during perceptual sequence learning. *J. Cogn. Neurosci.*, *28*(3), 402–417.

Altmann, G. T. M. (2017). Abstraction and generalization in statistical learning: Implications for the relationship between semantic types and episodic tokens. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *372*(1711).

Aly, M., Chen, J., Turk-Browne, N. B. & Hasson, U. (2018). Learning naturalistic temporal structure in the posterior medial network. *J. Cogn. Neurosci.*, *30*(9), 1345–1365.

Amano, K., Shibata, K., Kawato, M., Sasaki, Y. & Watanabe, T. (2016). Learning to associate orientation with color in early visual areas by associative decoded fMRI neurofeedback. *Curr. Biol.*, *26*(14), 1861–1866.

Amedi, A, Malach, R, Hendler, T, Peled, S & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway.

Amitay, S., Hawkey, D. & Moore, D. (2005). Auditory frequency discrimination learning is affected by stimulus variability. *Perception & Psychophysics*, *67*, 691–698. https://doi.org/10.3758/BF03193525

Aslin, R. N. (2017). Statistical learning: A powerful mechanism that operates by mere exposure. *Wiley Interdiscip. Rev. Cogn. Sci.*, *8*, 1–2. http://dx.doi.org/10.1002/wcs.1373

Aslin, R. N., Saffran, J. R. & Newport, E. L. (1998). Computation of conditional probability statistics by 8-Month-Old infants.

Astle, A. T., Li, R. W., Webb, B. S., Levi, D. M. & McGraw, P. V. (2013). A weber-like law for perceptual learning. *Sci. Rep.*, *3*.

Austerweil, J. L. & Griffiths, T. L. (2011). A rational model of the effects of distributional information on feature learning. *Cogn. Psychol.*, *63*(4), 173–209.

Baldauf, D & Desimone, R. (2014). Neural mechanisms of Object-Based attention. *Science*, *344*(6182), 424–427.

Ball, K. & Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science*, *218*, 697–698.

Ball, K & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Res.*, *27*(6), 953–965.

Banai, K., Ortiz, J., Oppenheimer, J. & Wright, B. (2010). Learning two things at once: Differential constraints on the acquisition and consolidation of perceptual learning. *Neuros-*

cience, *165*(2), 436–444. https://doi.org/https://doi.org/10.1016/j.neuroscience.2009.10.060

Barakat, B. K., Seitz, A. R. & Shams, L. (2013). The effect of statistical learning on internal stimulus representations: Predictable items are enhanced even when not predicted. *Cognition*, *129*(2), 205–211.

Batterink, L. & Paller, K. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31–45.

Baum, L. E. (1972). An inequality and associated maximization technique in statistical estimation of probabilistic functions of markov processes. *Inequalities*, *3*, 1–8.

Baylis, G. C. & Driver, J. (1993). Visual attention and objects: Evidence for hierarchical coding of location. *J. Exp. Psychol. Hum. Percept. Perform.*, *19*(3), 451–470.

Behrens, T. E. J., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Bejjanki, V. R., Beck, J. M., Lu, Z.-L. & Pouget, A. (2011). Perceptual learning as improved probabilistic inference in early sensory areas. *Nat. Neurosci.*, *14*(5), 642–648.

Bertamini, M. (2001). The importance of being convex: An advantage for convexity when judging position. *Perception*, *30*(11), 1295–1310.

Bertenthal, B. I. (1996). Origins and early development of perception, action, and representation. *Annu. Rev. Psychol.*, *47*, 431–459.

Bertenthal, B. I. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*, 1193–1216.

Bi, T., Chen, J., Zhou, T., He, Y. & Fang, F. (2014). Function and structure of human left fusiform cortex are closely associated with perceptual learning of faces. *Curr. Biol.*, *24*(2), 222–227.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychol. Rev.*, *94*(2), 115–147.

Bishop, C. M. (2006). *Pattern recognition and machine learning (information science and statistics)*. Springer-Verlag.

Bishop, C. M. (2016, August). *Pattern recognition and machine learning*. Springer.

Brady, T. F. & Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychol. Sci.*, *19*(7), 678–685.

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.*, *10*(4), 433–436.

Braun, D. A., Aertsen, A., Wolpert, D. M. & Mehring, C. (2004). Motor task variation induces structural learning. *Curr. Biol.*, *19*, 352–357.

Braun, D. A., Mehring, C. & Wolpert, D. M. (2010). Structure learning in action. *Behav. Brain Res.*, *206*, 157–165.

Brunel, N. & Nadal, J.-P. (1998). Mutual Information, Fisher Information, and Population Coding. *Neural Computation*, *10*(7), 1731–1757. https://doi.org/10.1162/089976698300017115

Bulf, H., Johnson, S. P. & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127–132.

Burton, G. J. (1981). Contrast discrimination by the human visual system. *Biol. Cybern.*, *40*(1), 27–38.

Carey, S. (2009a). The origin of concepts.

Carey, S. (2009b, May). *The origin of concepts*. Oxford University Press.

Castro, L., Wasserman, E. A. & Lauffer, M. (2018). Unsupervised learning of complex associations in an animal model. *Cognition*, *173*, 28–33.

Chang, D. H. F., Kourtzi, Z. & Welchman, A. E. (2013). Mechanisms for extracting a signal from noise as revealed through the specificity and generality of task training. *J. Neurosci.*, *33*(27), 10962–10971.

Chang, D. H. F., Mevorach, C., Kourtzi, Z. & Welchman, A. E. (2014). Training transfers the limits on perception from parietal to ventral cortex. *Curr. Biol.*, *24*(20), 2445–2450.

Chen, N., Cai, P., Zhou, T., Thompson, B. & Fang, F. (2016). Perceptual learning modifies the functional specializations of visual cortical areas. *Proc. Natl. Acad. Sci. U. S. A.*, *113*(20), 5724–5729.

Cohen, J. (2013, May). *Statistical power analysis for the behavioral sciences*. Routledge.

Cohen, Y., Daikhin, L. & Ahissar, M. (2013). Perceptual learning is specific to the trained structure of information. *J. Cogn. Neurosci.*, *25*(12), 2047–2060.

Cole, R. A. (1980). *Perception and production of fluent speech*. Routledge.

Cong, L.-J. & Zhang, J.-Y. (2014). Perceptual learning of contrast discrimination under roving: The role of semantic sequence in stimulus tagging. *Journal of Vision*, *14*(13), 1–1. https://doi.org/10.1167/14.13.1

contributors., W. (2018). *Wolpertinger*. Wikipedia, the Free Encyclopedia. [Accessed January 9, 2018]. https://en.wikipedia.org/w/index.php?title=Wolpertinger&oldid=833678893

Conway, C. M. & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J. Exp. Psychol. Learn. Mem. Cogn.*, *31*(1), 24–39.

Conway, C. M. & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological science*, *17*(10), 905–912.

Cox, D. & Hinkley, D. (1974). *Theoretical statistics (1st ed.)* Chapman; Hall/CRC.

Creel, S. C., Newport, E. L. & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences.

Crist, R. E., Kapadia, M. K., Westheimer, G & Gilbert, C. D. (1997). Perceptual learning of spatial localization: Specificity for orientation, position, and context. *J. Neurophysiol.*, *78*(6), 2889–2894.

Csibra, G & Gergely, G. Social learning and social cognition: The case for pedagogy (Y Munakata & M, Ed.). In: In *Processes of change in brain and cognitive Development.Attention and performance, XXI* (Y Munakata & M, Ed.). Ed. by Y Munakata & M. Vol. 21. Oxford University Press, 2006, pp. 249–274.

Dayan, P. & Abbott, L. F. (2005). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. The MIT Press.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–38. http://www.jstor.org/stable/2984875

Deneve, S. Bayesian inference in spiking neurons (L. Saul, Y. Weiss & L. Bottou, Eds.). In: In *Advances in neural information processing systems* (L. Saul, Y. Weiss & L. Bottou, Eds.). Ed. by Saul, L., Weiss, Y. & Bottou, L. *17*. MIT Press, 2005. https://proceedings.neurips.cc/paper/2004/file/cdd96eedd7f695f4d61802f8105ba2b0-Paper.pdf

DeValois, K. (1977). Spatial frequency adaptation can enhance contrast sensitivity. *Vision Research*, *17*, 1057–1065.

Devillez, H., Mollison, M. V., Hagen, S., Tanaka, J. W., Scott, L. S. & Curran, T. (2018). Color and spatial frequency differentially impact early stages of perceptual expertise training. *Neuropsychologia*.

Diaz, J. A., Queirazza, F. & Philiastides, M. G. (2017). Perceptual learning alters post-sensory processing in human decision-making. *Nature Human Behaviour*, *1*(2), 0035.

DiCarlo, J. J., Zoccolan, D. & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73(3)*, 415–434. http://doi.org/10.1016/j.neuron.2012.01.010

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*, 274–290. https://doi.org/10.1177/1745691611406920

Donovan, I. & Carrasco, M. (2018). Endogenous spatial attention during perceptual learning facilitates location transfer. *J. Vis.*, *18*(11), 7.

Donovan, I., Szpiro, S. & Carrasco, M. (2015). Exogenous attention facilitates location transfer of perceptual learning. *J. Vis.*, *15*(10), 11.

Dosher, B. & Lu, Z.-L. (2017). Visual perceptual learning and models. *Annual Review of Vision Science*, *3*(1), 343–363.

Dosher, B. A., Jeter, P., Liu, J. & Lu, Z.-L. (2013). An integrated reweighting theory of perceptual learning. *Proc. Natl. Acad. Sci. U. S. A.*, *110*(33), 13678–13683.

Dosher, B. A., Liu, J., Chu, W. & Lu, Z.-L. (2020). Roving: The causes of interference and re-enabled learning in multi-task visual training. *Journal of Vision*, *20*(6), 9–9. https://doi.org/10.1167/jov.20.6.9

Duncan, J. (1984). Selective attention and the organization of visual information. *J. Exp. Psychol. Gen.*, *113*(4), 501–517.

Echeveste, R., Aitchison, L., Hennequin, G. & Lengyel, M. (2020). Cortical-like dynamics in recurrent circuits optimized for sampling-based probabilistic inference. *Nature Neuroscience*, *23*(9), 1138–1149. https://doi.org/10.1038/s41593-020-0671-1

Egly, R, Driver, J & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *J. Exp. Psychol. Gen.*, *123*(2), 161–177.

Ellis, K., Morales, L., Sablé-Meyer, M., Solar-Lezama, A. & Tenenbaum, J. Learning libraries of subroutines for neurally–guided bayesian program induction (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi & R. Garnett, Eds.). In: In *Advances in neural information processing systems* (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi & R. Garnett, Eds.). Ed. by Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N. & Garnett, R. *31*. Curran Associates, Inc., 2018. https://proceedings.neurips.cc/paper/2018/file/7aa685b3b1dc1d6780bf36f7340078c9-Paper.pdf

Erickson, L. C. & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Dev. Rev.*, *37*, 66–108.

Fahle, M. (1997). Specificity of learning curvature, orientation, and vernier discriminations. *Vision Res.*, *37*(14), 1885–1895.

Fahle, M & Henke-Fahle, S. (1996). Interobserver variance in perceptual performance and learning. *Invest. Ophthalmol. Vis. Sci.*, *37*(5), 869–877.

Fahle, M & Morgan, M. (1996). No transfer of perceptual learning between similar stimuli in the same retinal position. *Curr. Biol.*, *6*(3), 292–297.

Fahle, M. & Poggio, T. (2002). *Perceptual learning*. MIT Press. http://dx.doi.org/10.7551/mitpress/5295.001.0001.

Fahle, M., Poggio, T. & Poggio, T. A. (2002). *Perceptual learning*. MIT Press.

Fechner, G. T. (1999). *The elements of psychophysics*. Breitkopf und Hartel (Reprinted, Bristol: Thoemmes Press).

Feldman, J. (1997). Regularity-based perceptual grouping. *Comput. Intell.*, *13*(4), 582–623.

Feldman, J. (2000). Bias toward regular form in mental shape spaces. *J. Exp. Psychol. Hum. Percept. Perform.*, *26*(1), 152–165.

Feldman, J. (2003). What is a visual object? *Trends Cogn. Sci.*, *7*(6), 252–256.

Fendick, M. & Westheimer, G. (1980). Effects of practice and the separation of test targets on foveal and peripheral stereoacuity. *Vision Research*, *287*, 43–44.

Finn, A., Kharitonova, M., Holtby, N. & Sheridan, M. (2018). Prefrontal and hippocampal structure predict statistical learning ability in early childhood. *J. Cogn. Neurosci.*, *31*, 126–137. https://doi.org/doi:10.1162/jocn_a_01342

Fiorentini, A. & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature*, *287*(5777), 43–44.

Fiser, J & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol. Sci.*, *12*(6), 499–504.

Fiser, J. (2009). The other kind of perceptual learning. *Learn. Percept.*, *1*(1), 69–87.

Fiser, J. & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proc. Natl. Acad. Sci. U. S. A.*, *99*(24), 15822–15826.

Fiser, J. & Aslin, R. N. (2005). Encoding multielement scenes: Statistical learning of visual feature hierarchies. *J. Exp. Psychol. Gen.*, *134*(4), 521–537.

Fiser, J., Berkes, P., Orbán, G. & Lengyel, M. (2010). Statistically optimal perception and learning: From behavior to neural representations. *Trends Cogn. Sci.*, *14*(3), 119–130.

Fiser, J. & Lengyel, G. (2019). A common probabilistic framework for perceptual and statistical learning [Computational Neuroscience]. *Current Opinion in Neurobiology*, *58*, 218–228. https://doi.org/https://doi.org/10.1016/j.conb.2019.09.007

Flanagan, J. & Beltzner, M. (2000). Independence of perceptual and sensorimotor predictions in the size-weight illusion. *Nature Neuroscience*, *3*, 737–741. https://doi.org/10.1038/76701

Frost, R., Armstrong, B. C., Siegelman, N. & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends Cogn. Sci.*, *19*(3), 117–125.

Froyen, V., Feldman, J. & Singh, M. (2015). Bayesian hierarchical grouping: Perceptual grouping as mixture estimation. *Psychol. Rev.*, *122*(4), 575–597.

Fu, Y., Hospedales, T., Xiang, T., Fu, Z. & Gong, S. Transductive multi-view embedding for zero-shot recognition and annotation. In: *Computer vision – european conference on computer*. 2014, 584–599. https://doi.org/10.1007/978-3-319-10605-2_38.

Ganguli, D. & Simoncelli, E. P. (2014). Efficient Sensory Encoding and Bayesian Inference with Heterogeneous Neural Populations. *Neural Computation*, *26*(10), 2103–2134. https://doi.org/10.1162/NECO_a_00638

García-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: Asymptotic and small-sample properties. *Vision Res.*, *38*(12), 1861–1881.

Geisler, W. S., Perry, J. S., Super, B. J. & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res.*, *41*(6), 711–724.

Gershman, S. J. & Niv, Y. (2010). Learning latent structure: Carving nature at its joints. *Curr. Opin. Neurobiol.*, *20*(2), 251–256.

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*, *521*(7553), 452–459.

Ghazanfar, A. A. & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cogn. Sci.*, *10*(6), 278–285.

Gibson, E. J. (1967). *Principles of perceptual learning and development*. Appleton Century Crofts. http://dx.doi.org/10.7551/mitpress/5295.001.0001.

Gibson, J. J. (1979). The ecological approach to visual perception.

Gilbert, C., Sigman, M. & Crist, R. (2001). The neural basis of perceptual learning. *Neuron*, *31*, 681–697.

Gilks, W. & Spiegelhalter, D. (1995). Markov chain monte carlo in practice.

Glicksohn, A. & Cohen, A. (2013). The role of cross-modal associations in statistical learning. *Psychon. Bull. Rev.*, *20*(6), 1161–1169.

Goldwater, S., Griffiths, T. L. & Johnson, M. (2009). A bayesian framework for word segmentation: Exploring the effects of context. *Cognition*, *112*(1), 21–54.

Goltstein, P. M., Coffey, E. B. J., Roelfsema, P. R. & Pennartz, C. M. A. (2013). In vivo two-photon ca2+ imaging reveals selective reward effects on stimulus-specific assemblies in mouse visual cortex. *J. Neurosci.*, *33*(28), 11540–11555.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T. & Danks, D. (2004). A theory of causal learning in children: Causal maps and bayes nets. *Psychological Review*, *111*(1), 3–32. https://doi.org/10.1037/0033-295X.111.1.3

Green, C. S., Kattner, F., Siegel, M. H., Kersten, D. & Schrater, P. R. (2015). Differences in perceptual learning transfer as a function of training task. *J. Vis.*, *15*(10), 5.

Green, C. S., Shawn Green, C, Li, R. & Bavelier, D. (2010a). Perceptual learning during action video game playing. *Top. Cogn. Sci.*, *2*(2), 202–216.

Green, C. S., Shawn Green, C, Pouget, A. & Bavelier, D. (2010b). Improved probabilistic inference as a general learning mechanism with action video games. *Curr. Biol.*, *20*(17), 1573–1579.

Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley.

Griffiths, T., Chater, N., Kemp, C., Perfors, A. & T.B., T. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in cognitive sciences*, *14*(8), 357–364.

Gu, Y., Liu, S., Fetsch, C. R., Yang, Y., Fok, S., Sunkara, A., DeAngelis, G. C. & Angelaki, D. E. (2011). Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron*, *71*(4), 750–761.

Haefner, R. M., Berkes, P. & Fiser, J. (2016). Perceptual Decision-Making as probabilistic inference by neural sampling. *Neuron*, *90*(3), 649–660.

Hastie, T., Tibshirani, R. & Friedman, J. (2013, November). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media.

Heald, J. B., Lengyel, M. & Wolpert, D. M. (2020). Contextual inference underlies the learning of sensorimotor repertoires. *bioRxiv*. https://doi.org/10.1101/2020.11.23.394320

Herzog, M. H. & Manfred, F. (1997). The role of feedback in learning a vernier discrimination task. *Vision Res.*, *37*(15), 2133–2141.

Heydt, R. v. d., von der Heydt, R, Peterhans, E & Baumgartner, G. (1984). Illusory contours and cortical neuron responses.

Howard, I., Ingram, J. & Wolpert, D. (2009). A modular planar robotic manipulandum with end-point torque control. *Journal of Neuroscience Methods*, *181*, 199–211. https://doi.org/10.1016/j.jneumeth.2009.05.005

Hua, T., Bao, P., Huang, C.-B., Wang, Z., Xu, J., Zhou, Y. & Lu, Z.-L. (2010). Perceptual learning improves contrast sensitivity of V1 neurons in cats. *Curr. Biol.*, *20*(10), 887–894.

Hung, S.-C. & Seitz, A. R. (2014). Prolonged training at threshold promotes robust retinotopic specificity in perceptual learning. *J. Neurosci.*, *34*(25), 8423–8431.

Hunt, R. H. & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners.

Hussain, Z., Bennett, P. J. & Sekuler, A. B. (2012). Versatile perceptual learning of textures after variable exposures. *Vision Res.*, *61*, 89–94.

Jacobs, R. A. & Kruschke, J. K. (2011). Bayesian learning theory applied to human cognition. *Wiley Interdiscip. Rev. Cogn. Sci.*, *2*(1), 8–21.

Jazayeri, M. & Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nat. Neurosci.*, *9*, 690–696. https://doi.org/doi.org/10.1038/nn1691

Jehee, J., Ling, S., Swisher, J., van Bergen, R. & Tong, F. (2012). Perceptual learning selectively refines orientation representations in early visual cortex. *J. Neurosci.*, *32*, 16747–53a.

Jeter, P. E., Dosher, B. A., Liu, S.-H. & Lu, Z.-L. (2010). Specificity of perceptual learning increases with increased training. *Vision Res.*, *50*(19), 1928–1940.

Jeter, P. E., Dosher, B. A., Petrov, A. & Lu, Z.-L. (2009). Task precision at transfer determines specificity of perceptual learning. *J. Vis.*, *9*(3), 1.1–13.

Jiang, J., Summerfield, C. & Egner, T. (2016). Visual prediction error spreads across object features in human visual cortex. *J. Neurosci.*, *36*(50), 12746–12763.

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S. & Saul, L. K. An introduction to variational methods for graphical models. In: *Learning in graphical models*. Cambridge, MA: MIT Press, 1999, 105–162.

Julesz, B. (1971). *Foundation of cyclopean perception*.

Kahnt, T., Grueschow, M., Speck, O. & Haynes, J.-D. (2011). Perceptual learning and decision-making in human medial frontal cortex. *Neuron*, *70*(3), 549–559.

Kanizsa, G. (1979, September). *Organization in vision: Essays on gestalt perception*. Praeger Publishers.

Kaposvari, P., Kumar, S. & Vogels, R. (2018). Statistical learning signals in macaque inferior temporal cortex. *Cereb. Cortex*, *28*(1), 250–266.

Karlaftis, V. M., Wang, R., Shen, Y., Tino, P., Williams, G., Welchman, A. E. & Kourtzi, Z. (2018). White-Matter pathways for statistical learning of temporal structures. *eNeuro*, *5*(3).

Karmali, F., Chaudhuri, S. E., Yi, Y. & Merfeld, D. M. (2016). Determining thresholds using adaptive procedures and psychometric fits: Evaluating efficiency using theory, simulations, and human experiments. *Exp. Brain Res.*, *234*(3), 773–789.

Karni, A & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proc. Natl. Acad. Sci. U. S. A.*, *88*(11), 4966–4970.

Karni, A, Tanne, D, Rubenstein, B, Askenasy, J & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science*, *265*(5172), 679–682.

Karni, A. & Sagi, D. (1993). The time course of learning a visual skill. *Nature*, *365*(6443), 250–252.

Karuza, E. A., Emberson, L. L., Roser, M. E., Cole, D., Aslin, R. N. & Fiser, J. (2017). Neural signatures of spatial statistical learning: Characterizing the extraction of structure from complex visual scenes. *J. Cogn. Neurosci.*, *29*(12), 1963–1976.

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E. & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language*, *127*(1), 46–54.

Kattner, F., Cochrane, A., Cox, C. R., Gorman, T. E. & Green, C. S. (2017). Perceptual learning generalization from sequential perceptual training as a change in learning rate. *Curr. Biol.*, *27*(6), 840–846.

Kellman, P. J. & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cogn. Psychol.*, *23*(2), 141–221.

Kellman, P. J. & Spelke, E. S. (1983). Perception of partly occluded objects in infancy.

Kemp, C & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, *105*(31), 10687–10692.

Kersten, D., Mamassian, P. & Yuille, A. (2004). Object perception as bayesian inference. *Annu. Rev. Psychol.*, *55*(1), 271–304.

Kim, R., Seitz, A., Feenstra, H. & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neurosci. Lett.*, *461*(2), 145–149.

Kirkham, N. Z., Slemmer, J. A. & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–42.

Knill, D. C. & Pouget, A. (2004). The bayesian brain: The role of uncertainty in neural coding and computation. *Trends Neurosci.*, *27*(12), 712–719.

Koblinger, Fiser, J. & Lengyel, M. (2021). Representations of uncertainty: Where art thou? *Current Opinion in Behavioral Sciences*, *38*, 150–162. https://doi.org/https://doi.org/10.1016/j.cobeha.2021.03.009

Koelsch, S., Busch, T., Jentschke, S. & Rohrmeier, M. (2016). Under the hood of statistical learning: A statistical mmn reflects the magnitude of transitional probabilities in auditory sequences. *Sci. Rep.*, *6*, 9741.

Köver, H., Gill, K., Tseng, Y.-T. L. & Bao, S. (2013). Perceptual and neuronal boundary learned from higher-order stimulus probabilities. *J. Neurosci.*, *33*(8), 3699–3705.

Kuai, S.-G., Levi, D. & Kourtzi, Z. (2013). Learning optimizes decision templates in the human visual cortex. *Curr. Biol.*, *23*(18), 1799–1804.

Kuai, S.-G., Zhang, J.-Y., Klein, S. A., Levi, D. M. & Yu, C. (2005). The essential role of stimulus temporal patterning in enabling perceptual learning. *Nat. Neurosci.*, *8*(11), 1497–1499.

Kwon, M., Daptardar, S., Schrater, P. & Pitkow, X. (2020a). Inverse rational control with partially observable continuous nonlinear dynamics.

Kwon, M., Daptardar, S., Schrater, P. R. & Pitkow, X. Inverse rational control with partially observable continuous nonlinear dynamics (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan & H. Lin, Eds.). In: In *Advances in neural information processing systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan & H. Lin, Eds.). Ed. by Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F. & Lin, H. *33*. Curran Associates, Inc., 2020, 7898–7909. https://proceedings.neurips.cc/paper/2020/file/5a01f0597ac4bdf35c24846734ee9a76-Paper.pdf

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B. & Shams, L. (2007). Causal inference in multisensory perception. *PLOS ONE*, *2*(9), 1–10. https://doi.org/10.1371/journal.pone.0000943

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. *Behav. Brain Sci.*, *40*, e253.

Lakshminarasimhan, K. J., Petsalis, M., Park, H., DeAngelis, G. C., Pitkow, X. & Angelaki, D. E. (2018). A dynamic bayesian observer model reveals origins of bias in visual path integration. *Neuron*, *99*(1), 194–206.e5. https://doi.org/https://doi.org/10.1016/j.neuron.2018.05.040

Lampert, C., Nickisch, H. & S., H. Learning to detect unseen object classes by between-class attribute transfer. In: *Ieee conference on computer vision and pattern recognition*. 2009, 951–958. https://doi.org/10.1109/cvprw.2009.5206594.

Law, C.-T. & Gold, J. I. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nat. Neurosci.*, *11*(4), 505–513.

Law, C.-T. & Gold, J. I. (2009). Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nat. Neurosci.*, *12*(5), 655–663.

Law, C.-T. & Gold, J. I. (2010). Shared mechanisms of perceptual learning and decision making. *Top. Cogn. Sci.*, *2*(2), 226–238.

Lee, H., Mozer, M. C., Kramer, A. F. & Vecera, S. P. (2012). Object-based control of attention is sensitive to recent experience. *J. Exp. Psychol. Hum. Percept. Perform.*, *38*(2), 314–325.

Legge, G. E. (1981). A power law for contrast discrimination. *Vision Res.*, *21*(4), 457–467.

Lehiste, I. (1970). *Suprasegmentals*. MIT Press.

LeMessurier, A. M. & Feldman, D. E. (2018). Plasticity of population coding in primary sensory cortex. *Curr. Opin. Neurobiol.*, *53*, 50–56.

Lengyel, G. & Fiser, J. (2019). The relationship between initial threshold, learning, and generalization in perceptual learning. *Journal of Vision*, *19*(4), 28–28. https://doi.org/10.1167/19.4.28

Lengyel, G., Nagy, M. & Fiser, J. (2021). Statistically defined visual chunks engage object-based attention. *Nature Communications*, *12*(1), 272. https://doi.org/10.1038/s41467-020-20589-z

Lengyel, G., Žalalytė, G., Pantelides, A., Ingram, J. N., Fiser, J., Lengyel, M. & Wolpert, D. M. (2019). Unimodal statistical learning produces multimodal object-like representations (J. Diedrichsen, M. J. Frank, M. Landy & S. J. Gershman, Eds.). *eLife*, *8*, e43942. https://doi.org/10.7554/eLife.43942

Leslie, A. M., Xu, F., Tremoulet, P. D. & Scholl, B. J. (1998). Indexing and the object concept: Developing 'what' and 'where' systems. *Trends Cogn. Sci.*, *2*(1), 10–18.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.*, *49*(2), Suppl 2:467+.

Li, W., Piëch, V. & Gilbert, C. D. (2004). Perceptual learning and top-down influences in primary visual cortex. *Nat. Neurosci.*, *7*(6), 651–657.

Liu, L. D. & Pack, C. C. (2017). The contribution of area MT to visual motion perception depends on training. *Neuron*, *95*(2), 436–446.e3.

Liu, Z, Jacobs, D. W. & Basri, R. (1999). The role of convexity in perceptual completion: Beyond good continuation. *Vision Res.*, *39*(25), 4244–4257.

Lu, Z.-L., Liu, J. & Dosher, B. A. (2010). Modeling mechanisms of perceptual learning with augmented hebbian re-weighting [Perceptual Learning Part II]. *Vision Research*, *50*(4), 375–390. https://doi.org/https://doi.org/10.1016/j.visres.2009.08.027

Luck, S. J. & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.

Luo, Y. & Zhao, J. (2018). Statistical learning creates novel object associations via transitive relations. *Psychol. Sci.*, *29*(8), 1207–1220.

Ly, A., Verhagen, J. & Wagenmakers, E.-J. (2016). Harold jeffreys's default bayes factor hypothesis tests: Explanation, extension, and application in psychology. *J. Math. Psychol.*, *72*, 19–32.

Ma, W. J., Beck, J. M., Latham, P. E. & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–1438. https://doi.org/10.1038/nn1790

Mach, E. (1861). Über das sehen von lagen und winkeln durch die bewegung des auges. *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften*, *43*(2), 215–224.

Manassi, M., Sayim, B. & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *J. Vis.*, *12*(10), 13.

Maniglia, M. & Seitz, A. R. (2018). Towards a whole brain model of perceptual learning. *Curr Opin Behav Sci*, *20*, 47–55.

Mansfield, R. J. W. (1974). Neural basis of orientation perception in primate vision. *Science*, *186*(4169), 1133–1135.

Marcus, G. F., Vijayan, S, Bandi Rao, S & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77–80.

Mareschal, D. & French, R. M. (2017). TRACX2: A connectionist autoencoder using graded chunks to model infant visual statistical learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *372*(1711).

Marr, D. (1982, January). *Vision: A computational investigation into the human representation and processing of visual information*.

McCullagh, P. & Nelder, J. (1989). *Generalized linear models. second edition*. CRC Press.

Meyer, T. & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proc. Natl. Acad. Sci. U. S. A.*, *108*(48), 19401–19406.

Michel, M. M. & Jacobs, R. A. (2007). Parameter learning but not structure learning: A bayesian network model of constraints on early perceptual learning. *J. Vis.*, *7*(1), 4.

Mikellidou, K., Cicchini, G. M., Thompson, P. G. & Burr, D. C. (2015). The oblique effect is both allocentric and egocentric. *J. Vis.*, *15*(8), 24.

Miyamoto, D, Hirai, D, Fung, C. C. A., Inutsuka, A, Odagawa, M, Suzuki, T, Boehringer, R, Adaikkan, C, Matsubara, C, Matsuki, N, Fukai, T, McHugh, T. J., Yamanaka, A & Murayama, M. (2016). Top-down cortical input during NREM sleep consolidates perceptual memory. *Science*, *352*(6291), 1315–1318.

Moore, C. M., Yantis, S. & Vaughan, B. (1998). Object-Based visual selection: Evidence from perceptual completion. *Psychol. Sci.*, *9*(2), 104–110.

Morey, R. D. & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychol. Methods*, *16*(4), 406–419.

Murphy, K. P. (2012, August). *Machine learning: A probabilistic perspective*. MIT Press.

Musz, E., Weber, M. J. & Thompson-Schill, S. L. (2014). Visual statistical learning is not reliably modulated by selective attention to isolated events. *Atten. Percept. Psychophys.*, *77*(1), 78–96.

Nagy, D. G., Török, B. & Orbán, G. (2020). Optimal forgetting: Semantic compression of episodic memories. *PLOS Computational Biology*, *16*(10), 1–28. https://doi.org/10.1371/journal.pcbi.1008367

Nahum, M., Nelken, I. & Ahissar, M. (2010). Stimulus uncertainty and perceptual learning: Similar principles govern auditory and visual learning [Perceptual Learning Part II]. *Vision Research*, *50*(4), 391–401. https://doi.org/https://doi.org/10.1016/j.visres.2009.09.004

Needham, A. (1997). Object segregation in 8-month-old infants. *Cognition*, *62*(2), 121–149.

Needham, A. & Baillargeon, R. (1998). Effects of prior experience on 4.5-month old infants' object segregation. *Infant Behav. Dev.*, *21*(1), 1–24.

Needham, A. & Modi, A. Infants' use of prior experiences with objects in object segregation: Implications for object recognition in infancy. In: *Advances in child development and behavior*. 1999, pp. 99–133.

Neil, M. H., Huszár, F., Mohammad, M. G., Orbán, G., Daniel, M. W. & Lengyel, M. (2013). Cognitive tomography reveals complex, task-independent mental representations. *Current Biology*, *23*(21), 2169 –2175. https://doi.org/https://doi.org/10.1016/j.cub.2013.09.012

Nemeth, D., Janacsek, K., Londe, Z., Ullman, M. T., Howard, D. V. & Howard, J. H. (2009). Sleep has no critical role in implicit motor sequence learning in young and old adults. *Exp. Brain Res.*, *201*(2), 351–358.

Newport, E. L. (2016). Statistical language learning: Computational, maturational, and linguistic constraints. *Lang. Cogn.*, *8*(03), 447–461.

Ni, A. M., Ruff, D. A., Alberts, J. J., Symmonds, J & Cohen, M. R. (2018). Learning and attention reveal a general relationship between population activity and behavior. *Science*, *359*(6374), 463–465.

O'Craven, K. M., Downing, P. E. & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*(6753), 584–587.

Olshausen, B. A. & Field, J. F. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.

Ongchoco, J., Uddenberg, S. & Chun, M. (2016). Statistical learning of movement. *J. Vis.*, *16*(12), 1079.

Orban, G. A., Vandenbussche, E & Vogels, R. (1984). Human orientation discrimination tested with long stimuli. *Vision Res.*, *24*(2), 121–128.

Orbán, G., Berkes, P., Fiser, J. & Lengyel, M. (2016). Neural variability and Sampling-Based probabilistic representations in the visual cortex. *Neuron*, *92*(2), 530–543.

Orbán, G., Fiser, J., Aslin, R. N. & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proc. Natl. Acad. Sci. U. S. A.*, *105*(7), 2745–2750.

O'Toole, A. J. & Kersten, D. J. (1992). Learning to see random-dot stereograms. *Perception*, *21*(2), 227–243.

Otsuka, S. & Saiki, J. (2016). Gift from statistical learning: Visual statistical learning enhances memory for sequence elements and impairs memory for items that disrupt regularities. *Cognition*, *147*, 113–126.

Otto, T. U., Herzog, M. H., Fahle, M. & Zhaoping, L. (2006). Perceptual learning with spatial uncertainties. *Vision Research*, *46*(19), 3223–3233. https://doi.org/https://doi.org/10.1016/j.visres.2006.03.021

Palmer, S. E. Perceptual organization in vision. In: *Stevens' handbook of experimental psychology*. 2002.

Palmer, S. E. & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychon. Bull. Rev.*, *1*(1), 29–55.

Parker, S. T. & Gibson, K. R. (1977). Object manipulation, tool use and sensorimotor intelligence as feeding adaptations in cebus monkeys and great apes. *J. Hum. Evol.*, *6*(7), 623–641.

Parkosadze, K., Otto, T. U., Malania, M., Kezeli, A. & Herzog, M. H. (2008). Perceptual learning of bisection stimuli under roving: Slow and largely specific. *Journal of Vision*, *8*(1), 5–5. https://doi.org/10.1167/8.1.5

Pascual-Leone, A. & Hamilton, R. (2001). The metamodal organization of the brain. *Progress in Brain Research*, *134*, 427.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat. Vis.*, *10*(4), 437–442.

Peña, M., Bonatti, L. L., Nespor, M. & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*(5593), 604–607.

Perruchet, P. & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*, 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Perruchet, P. (2018). What mechanisms underlie implicit statistical learning? transitional probabilities versus chunks in language learning. *Top. Cogn. Sci.*

Perruchet, P. (2019). What mechanisms underlie implicit statistical learning? transitional probabilities versus chunks in language learning. *Top. Cogn. Sci.*, *11*(3), 520–535.

Peterson, M. A. (1994). Object recognition processes can and do operate before Figure–Ground organization.

Petrov, A. A., Dosher, B. A. & Lu, Z.-L. (2005). The dynamics of perceptual learning: An incremental reweighting model. *Psychol. Rev.*, *112*(4), 715–743.

Petrov, A. A., Dosher, B. A. & Lu, Z.-L. (2006). Perceptual learning without feedback in non-stationary contexts: Data and model. *Vision Res.*, *46*(19), 3177–3197.

Piazza, E. A., Denison, R. N. & Silver, M. A. (2018). Recent cross-modal statistical learning influences visual perceptual selection. *J. Vis.*, *18*(3), 1.

Piëch, V., Li, W., Reeke, G. N. & Gilbert, C. D. (2013). Network model of top-down influences on local gain and contextual interactions in visual cortex. *Proc. Natl. Acad. Sci. U. S. A.*, *110*(43), E4108–17.

Pinker, S. (1998). Words and rules. *Lingua*, *106*, 219–242.

Piray, P. & Daw, N. D. (2020). A simple model for learning in volatile environments (A. Soltani, Ed.). *PLOS Computational Biology*, *16*(7), e1007963. https://doi.org/10.1371/journal.pcbi.1007963

Pizlo, Z., Salach-Golyska, M. & Rosenfeld, A. (1997). Curve detection in a noisy image. *Vision Res.*, *37*(9), 1217–1241.

Polat, U., Schor, C., Tong, J.-L., Zomet, A., Lev, M., Yehezkel, O., Sterkin, A. & Levi, D. M. (2012). Training the brain to overcome the effect of aging on the human eye. *Sci. Rep.*, *2*, 278.

Pomerantz, J. R., Sager, L. C. & Stoever, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *J. Exp. Psychol. Hum. Percept. Perform.*, *3*(3), 422–435.

Pouget, A., Beck, J. M., Ma, W. J. & Latham, P. E. (2013). Probabilistic brains: Knowns and unknowns. *Nat. Neurosci.*, *16*(9), 1170–1178.

Pouget, A., Dayan, P. & Zemel, R. S. (2003). Inference and computation with population codes. *Annu. Rev. Neurosci.*, *26*, 381–410.

Quinn, P. C. & Bhatt, R. S. (2005). Learning perceptual organization in infancy. *Psychol. Sci.*, *16*(7), 511–515.

Rabiner, L. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286. https://doi.org/10.1109/5.18626

Ramachandran, S., Meyer, T. & Olson, C. R. (2016). Prediction suppression in monkey inferotemporal cortex depends on the conditional probability between images. *J. Neurophysiol.*, *115*(1), 355–362.

Ramachandran, V. & Braddick, O. (1973). Orientation-specific learning in stereopsis. *Perception*, *2*, 371–376.

Rao, R. P. N., Olshausen, B. A. & Lewicki, M. S. (2002, March). *Probabilistic models of the brain: Perception and neural function*. MIT Press.

Rasmussen, C. E. (2004). Gaussian processes in machine learning. In O. Bousquet, U. von Luxburg & G. Rätsch (Eds.), *Advanced lectures on machine learning: Ml summer schools 2003, canberra, australia, february 2 - 14, 2003, tübingen, germany, august 4 - 16, 2003, revised lectures* (pp. 63–71). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-28650-9_4

Regan, D & Price, P. (1986). Periodicity in orientation discrimination and the unconfounding of visual information. *Vision Res.*, *26*(8), 1299–1302.

Rescorla, R. A. (1967). Pavlovian conditioning and its proper control procedures. *Psychol. Rev.*, *74*(1), 71–80.

Riesenhuber, M. & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.*, *2*(11), 1019–1025.

Roelfsema, P., van Ooyen, A. & Watanabe, T. (2010). Perceptual learning rules based on reinforcers and attention. *Trends Cogn. Sci.*, *14*, 64–71.

Rosa-Salva, O., Fiser, J., Versace, E., Dolci, C., Chehaimi, S., Santolin, C. & Vallortigara, G. (2018). Spontaneous learning of visual structures in domestic chicks. *Animals (Basel)*, *8*(8).

Rosenthal, C. R., Kennard, C. & Soto, D. (2010). Visuospatial sequence learning without seeing. *PLoS ONE*, *5*(7), e11906.

Rouder, J. N., Speckman, P. L., Dongchu, S., Morey, R. D. & Geoffrey, I. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.*, *16*(2), 225–237.

Ruthotto, L. & Haber, E. (2021). An introduction to deep generative modeling.

Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996). Statistical learning by 8-Month-Old infants. *Science*, *274*(5294), 1926–1928.

Saffran, J. R. & Kirkham, N. Z. (2018). Infant statistical learning. *Annu. Rev. Psychol.*, *69(1)*, 181–203. http://dx.doi.org/10.1146/annurevpsych-122216-011805

Saffran, J. R., Pollak, S. D., Seibel, R. L. & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition*, *105*(3), 669–680.

Sagi, D. (1994). Perceptual learning: Learning to see. *Curr. Opin. Neurobiol.*, *4(2)*, 195–199.

Sanayei, M., Chen, X., Chicharro, D., Distler, C., Panzeri, S. & Thiele, A. (2018). Perceptual learning of fine contrast discrimination changes neuronal tuning and population coding in macaque V4. *Nat. Commun.*, *9*(1), 4238.

Santolin, C., Rosa-Salva, O., Vallortigara, G. & Regolin, L. (2016). Unsupervised statistical learning in newly hatched chicks. *Curr. Biol.*, *26*(23), R1218–R1220.

Santolin, C. & Saffran, J. R. (2018). Constraints on statistical learning across species. *Trends Cogn. Sci.*, *22*(1), 52–63.

Schapiro, A., Gregory, E., Landau, B., McCloskey, M. & Turk-Browne, N. (2014). The necessity of the medial temporal lobe for statistical learning. *J. Cogn. Neurosci.*, *26*, 1736–1747.

Schapiro, A. C., Kustner, L. V. & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology*, *22*(17), 1622–1627.

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B. & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nat. Neurosci.*, *16*(4), 486–492.

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M. & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *372*(1711).

Schofield, A. J. (2000). What does second-order vision see in an image? *Perception*, *29*(9), 1071–1086.

Schofield, A. J., Rock, P. B., Sun, P., Jiang, X. & Georgeson, M. A. (2010). What is second-order vision for? discriminating illumination versus material changes. *J. Vis.*, *10*(9), 2.

Schoups, A, Vogels, R, Qian, N & Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature*, *412*(6846), 549–553.

Schoups, A. A., Vogels, R & Orban, G. A. (1995). Human perceptual learning in identifying the oblique orientation: Retinotopy, orientation specificity and monocularity. *J. Physiol.*, *483 (Pt 3)*, 797–810.

Schwabe, L. (2005). Adaptivity of tuning functions in a generic recurrent network model of a cortical hypercolumn. *Journal of Neuroscience*, *25*(13), 3323–3332.

Seitz, A. R. & Watanabe, T. (2003). Psychophysics: Is subliminal learning really passive? *Nature*, *422*(6927), 36.

Seriès, P., Stocker, A. A. & Simoncelli, E. P. (2009). Is the Homunculus "Aware" of Sensory Adaptation? *Neural Computation*, *21*(12), 3271–3304. https://doi.org/10.1162/neco.2009.09-08-869

Serre, T., Oliva, A. & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, *104(15)*, 6424–6429.

Seung, H. S. & Sompolinsky, H. (1993). Simple models for reading neuronal population codes. *Proceedings of the National Academy of Sciences*, *90*(22), 10749–10753. https://doi.org/10.1073/pnas.90.22.10749

Shanks, D. R. (2010). Learning: From association to cognition. *Annu. Rev. Psychol.*, *61*(1), 273–301.

Shibata, K., Sagi, D. & Watanabe, T. (2014). Two-stage model in perceptual learning: Toward a unified theory. *Ann. N. Y. Acad. Sci.*, *1316*, 18–28.

Shiu, L. P. & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Percept. Psychophys.*, *52*(5), 582–588.

Shomstein, S. & Yantis, S. (2004). Configural and contextual prioritization in object-based attention. *Psychon. Bull. Rev.*, *11*(2), 247–253.

Simor, P., Zavecz, Z., Horváth, K., Éltető, N., Török, C., Pesthy, O., Gombos, F., Janacsek, K. & Nemeth, D. (2018). Deconstructing procedural memory: Different learning trajectories and consolidation of sequence and statistical learning. *Front. Psychol.*, *9*, 2708.

Slone, L. K. & Johnson, S. P. (2018). When learning goes beyond statistics: Infants represent visual sequences in terms of chunks. *Cognition*, *178*, 92–102.

Smits, J. T. & Vos, P. G. (1987). The perception of continuous curves in dot stimuli. *Perception*, *16*(1), 121–131.

Solomon, J. A. & Tyler, C. W. (2017). Improvement of contrast sensitivity with practice is not compatible with a sensory threshold account. *J. Opt. Soc. Am.*, *34*(6), 870.

Sotiropoulos, G., Seitz, A. R. & Seriès, P. (2011). Changing expectations about speed alters perceived motion direction. *Curr. Biol.*, *21*(21), R883–4.

Spang, K, Grimsen, C, Herzog, M. H. & Fahle, M. (2010). Orientation specificity of learning vernier discriminations. *Vision Res.*, *50*(4), 479–485.

Spelke, E. S. (1990). Principles of object perception. *Cogn. Sci.*, *14*, 29–56.

Stansbury, D., Naselaris, T. & Gallant, J. (2013). Natural scene statistics account for the representation of scene categories in human visual cortex. *Neuron*, *79*, 1025–1034.

Stevens, S. S. (1957). On the psychophysical law. *Psychol. Rev.*, *64*(3), 153–181.

Stocker, A. A. & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature neuroscience*, *9*(4), 578–585. https://doi.org/10.1038/nn1669

Streiner, D. L. (2003). Unicorns do exist: A tutorial on "proving" the null hypothesis. *Can. J. Psychiatry*, *48*(11), 756–761.

Streri, A & Spelke, E. S. (1988). Haptic perception of objects in infancy.

Sun, Y. & Fisher, R. (2003). Object-based visual attention for computer vision.

Talluri, B., Hung, S.C., Seitz, A. & Seriès, P. (2015). Confidence-based integrated reweighting model of task-difficulty explains location-based specificity in perceptual learning. *J. Vis.*, *15*, 17.

Tan, Q., Wang, Z., Sasaki, Y. & Watanabe, T. (2019). Category-Induced transfer of visual perceptual learning. *Curr. Biol.*, *29*(8), 1374–1378.e3.

Tanaka, J. W., Curran, T. & Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychol. Sci.*, *16*(2), 145–151.

Tartaglia, E. M., Aberg, K. C. & Herzog, M. H. (2009a). Modeling perceptual learning: Why mice do not play backgammon. *Learning & Perception*, *1*(1), 155 –163. https://doi.org/10.1556/lp.1.2009.1.12

Tartaglia, E. M., Aberg, K. C. & Herzog, M. H. (2009b). Perceptual learning and roving: Stimulus types and overlapping neural populations. *Vision Research*, *49*(11), 1420–1427. https://doi.org/https://doi.org/10.1016/j.visres.2009.02.013

Tartaglia, E. M., Bamert, L., Mast, F. W. & Herzog, M. H. (2009c). Human perceptual learning by mental imagery. *Curr. Biol.*, *19*(24), 2081–2085.

Teich, A. F. & Qian, N. (2003). Learning and adaptation in a recurrent model of V1 orientation selectivity. *Journal of Neurophysiology*, *89*(4), 2086–2100.

Tenenbaum, J. B., Kemp, C, Griffiths, T. L. & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.

Tenenbaum, J. B., Griffiths, T. L. & Kemp, C. (2006). Theory-based bayesian models of inductive learning and reasoning. *Trends Cogn. Sci.*, *10*(7), 309–318.

Thiessen, E. D. (2011). Domain general constraints on statistical learning. *Child Dev.*, *82*(2), 462–470.

Toro, J. M. & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Percept. Psychophys.*, *67*(5), 867–875.

Tsodyks, M. & Gilbert, C. (2004). Neural networks and perceptual learning. *Nature*, *431*, 775–781.

Turk-Browne, N. B., Isola, P. J., Scholl, B. J. & Treat, T. A. (2008). Multidimensional visual statistical learning. *J. Exp. Psychol. Learn. Mem. Cogn.*, *34*(2), 399–407.

Turk-Browne, N. B., Jungé, J. & Scholl, B. J. (2005). The automaticity of visual statistical learning. *J. Exp. Psychol. Gen.*, *134*(4), 552–564.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M. & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *J. Cogn. Neurosci.*, *21*(10), 1934–1945.

Vasudeva Raju, R. & Pitkow, Z. Inference by reparameterization in neural population codes (D. Lee, M. Sugiyama, U. Luxburg, I. Guyon & R. Garnett, Eds.). In: In *Advances in neural information processing systems* (D. Lee, M. Sugiyama, U. Luxburg, I. Guyon & R. Garnett, Eds.). Ed. by Lee, D., Sugiyama, M., Luxburg, U., Guyon, I. & Garnett, R. *29*. Curran Associates, Inc., 2016. https://proceedings.neurips.cc/paper/2016/file/a26398dca6f47b49876cbaffbc9954f9-Paper.pdf

Vecera, S. P. (1994). Grouped locations and object-based attention: Comment on egly, driver, and rafal (1994). *J. Exp. Psychol. Gen.*, *123*(3), 316–320.

Vecera, S. P., Behrmann, M & McGoldrick, J. (2000). Selective attention to the parts of an object. *Psychon. Bull. Rev.*, *7*(2), 301–308.

Venzon, D. & Moolgavkar, S. (1988). A method for computing profile-likelihood-based confidence intervals. *Applied Statistics*, *37*, 87. https://doi.org/10.2307/2347496

Vértes, E. & Sahani, M. Flexible and accurate inference and learning for deep generative models (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi & R. Garnett, Eds.). In: In *Advances in neural information processing systems* (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi & R. Garnett, Eds.). Ed. by Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N. & Garnett, R. *31*. Curran Associates, Inc., 2018. https://proceedings.neurips.cc/paper/2018/file/955cb567b6e38f4c6b3f28cc857fc38c-Paper.pdf

Vickery, T. J. & Jiang, Y. V. (2009). Associative grouping: Perceptual grouping of shapes by association. *Atten. Percept. Psychophys.*, *71*(4), 896–909.

Vogels, R. (2010). Mechanisms of visual perceptual learning in macaque visual cortex. *Top. Cogn. Sci.*, *2*, 239–250.

Vogels, R. & Orban, G. A. (1985). The effect of practice on the oblique effect in line orientation judgments. *Vision Research*, *25*, 1679–1687.

Wang, R, Wang, J, Y. Zhang, J, Y. Xie, X, X. Yang, Y, H. Luo, S, Yu, C & Li, W. (2016). Perceptual learning at a conceptual level. *Journal of Neuroscience*, *36*(7), 2238–2246.

Wang, R., Shen, Y., Tino, P., Welchman, A. E. & Kourtzi, Z. (2017). Learning predictive statistics: Strategies and brain mechanisms. *J. Neurosci.*, *37*(35), 8412–8427.

Wang, R., Zhang, J.-Y., Klein, S. A., Levi, D. M. & Yu, C. (2014). Vernier perceptual learning transfers to completely untrained retinal locations after double training: A "piggybacking" effect. *J. Vis.*, *14*(13), 12.

Watanabe, T. & Sasaki, Y. (2015). Perceptual learning: Toward a comprehensive theory. *Annu. Rev. Psychol.*, *66*(1), 197–221.

Watson, A. & Fitzhugh, A. (1990). The method of constant stimuli is inefficient. *Perception & Psychophysics*, *47*, 87–91. https://doi.org/doi.org/10.3758/BF03208169

Weber, E. H. (1834). *De pulsu, resorptione, auditu et tactu: Annotationes anatomicae et physiologicae, auctore*.

Wei, X.-X. & Stocker, A. A. (2015). A bayesian observer model constrained by efficient coding can explain 'anti-bayesian' percepts. *Nature Neuroscience*, *18*(10), 1509–1517. https://doi.org/10.1038/nn.4105

Weiss, Y., Edelman, S. & Fahle, M. (1993). Models of perceptual learning in vernier hyperacuity. *Neural Comput.*, *5*(5), 695–718.

Wu, Z., Schrater, P. & Pitkow, X. (2018a). Inverse pomdp: Inferring what you think from what you do. *ArXiv*, *abs/1805.09864*.

Wu, Z., Schrater, P. & Pitkow, X. (2018b). Inverse rational control: Inferring what you think from how you forage. *arXiv: Learning*.

Xiao, L.-Q., Zhang, J.-Y., Wang, R., Klein, S. A., Levi, D. M. & Yu, C. (2008). Complete transfer of perceptual learning across retinal locations enabled by double training. *Curr. Biol.*, *18*(24), 1922–1926.

Xiong, Y.-Z., Zhang, J.-Y. & Yu, C. (2016). Bottom-up and top-down influences at untrained conditions determine perceptual learning specificity and transfer. *Elife*, *5*.

Xu, F. & Tenenbaum, J. B. (2007). Word learning as bayesian inference. *Psychological review*, *114*(2), 245–272. https://doi.org/10.1037/0033-295X.114.2.245

Yan, Y., Rasch, M. J., Chen, M., Xiang, X., Huang, M., Wu, S. & Li, W. (2014). Perceptual training continuously refines neuronal population codes in primary visual cortex. *Nat. Neurosci.*, *17*(10), 1380–1387.

Yang, T. & Maunsell, J. H. R. (2004). The effect of perceptual learning on neuronal responses in monkey visual area V4. *J. Neurosci.*, *24*(7), 1617–1626.

Yehezkel, O., Sterkin, A., Lev, M., Levi, D. M. & Polat, U. (2016). Gains following perceptual learning are closely linked to the initial visual acuity. *Scientific Reports*, *6*, 25188.

Yildirim, I. & Jacobs, R. (2012). A rational analysis of the acquisition of multisensory representations. *Cognitive Science*, *36*, 305–332. https://doi.org/10.1111/j.1551-6709.2011.01216.x

Yildirim, I. & Jacobs, R. (2013). Transfer of object category knowledge across visual and haptic modalities: Experimental and computational studies. *Cognition*, *126*, 135–148. https://doi.org/10.1016/j.cognition.2012.08.005

Yu, C., Klein, S. A. & Levi, D. M. (2004). Perceptual learning in contrast discrimination and the (minimal) role of context. *J. Vis.*, *4*(3), 169–182.

Yu, Q., Zhang, P., Qiu, J. & Fang, F. (2016). Perceptual learning of contrast detection in the human lateral geniculate nucleus. *Curr. Biol.*, *26*(23), 3176–3182.

Yu, R. Q. & Zhao, J. (2018). Implicit updating of object representation via temporal associations. *Cognition*, *181*, 127–134.

Zhang, J.-Y., Kuai, S.-G., Xiao, L.-Q., Klein, S. A., Levi, D. M. & Yu, C. (2008). Stimulus coding rules for perceptual learning. *PLOS Biology*, *6*(8), 1–10. https://doi.org/10.1371/journal.pbio.0060197

Zhang, J.-Y., Zhang, G.-L., Xiao, L.-Q., Klein, S. A., Levi, D. M. & Yu, C. (2010). Rule-based learning explains visual perceptual learning and its specificity and transfer. *J. Neurosci.*, *30*(37), 12323–12328.

Zhao, J., Al-Aidroos, N. & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychol. Sci.*, *24*(5), 667–677.

Zhao, J., Ngo, N., McKendrick, R. & Turk-Browne, N. B. (2011). Mutual interference between statistical summary perception and statistical learning. *Psychol. Sci.*, *22*(9), 1212–1219.

Zhao, J. & Yu, R. Q. (2016). Statistical regularities reduce perceived numerosity. *Cognition*, *146*, 217–222.

Zhao, L., Cosman, J. D., Vatterott, D. B., Gupta, P. & Vecera, S. P. (2014). Visual statistical learning can drive object-based attentional selection. *Atten. Percept. Psychophys.*, *76*(8), 2240–2248.

Zhou, H, Friedman, H. S. & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *J. Neurosci.*, *20*(17), 6594–6611.