Annu. Rev. Vis. Sci. 2022. 8:X–X
https://doi.org/10.1146/annurev-vision-100720-103343

Statistical Learning in Vision

József Fiser and Gábor Lengyel

Department of Cognitive Science, Center for Cognitive Computation, Central European University, Vienna 1100, Austria; email: fiserj@ceu.edu

[perceptual learning, rule learning, probabilistic computation, hierarchical Bayesian modeling]

Abstract

Vision and learning have long been considered to be two areas of research linked only distantly. However, recent developments in vision research have changed the conceptual definition of vision from a signal-evaluating process to a goal-oriented interpreting process, and this shift binds learning, together with the resulting internal representations, intimately to vision. In this review, we consider various types of learning (perceptual, statistical, and rule/abstract) associated with vision in the past decades and argue that they represent differently specialized versions of the fundamental learning process, which must be captured in its entirety when applied to complex visual processes. We show why the generalized version of statistical learning can provide the appropriate setup for such a unified treatment of learning in vision, what computational framework best accommodates this kind of statistical learning, and what plausible neural scheme could feasibly implement this framework. Finally, we list the challenges that the field of statistical learning faces in fulfilling the promise of being the right vehicle for advancing our understanding of vision in its entirety.

1. INTRODUCTION

Modern vision research has evolved from its early days, when spatial vision was its dominant area of study (DeValois & DeValois 1990, Graham 1989), with a heavy focus on the behavioral and neural details of low-level visual processing of spatial frequency, orientation, binocularity, or motion, into a field in which contextual information, perceptual biases, short- and long-term memory, the task at hand, and cognitive factors play as much of a role in understanding the process as does the fine quantification of the incoming sensory input (Cicchini et al. 2021, de Lange et al. 2018, Feldman 1997, Grosof et al. 1993, Maunsell 2015, Murray 2021, Palmer & Rock 1994, Sotiropoulos et al. 2011, Wagemans et al. 2012, von der Heydt et al. 1984). One direct ramification of this conceptual shift is a change in the status of learning and the internal representations that it creates in the context of visual processing. When the information of interest was defined as the orientation of a small edge segment or the brightness of a patch in a particular region of the visual field, it was possible to conceptualize the task of vision as figuring out a scalar value of a particular single dimension conveyed through the light falling on a part of the retina that can be handled by an appropriately tuned feature detector (DiCarlo et al. 2012, Marr 1982, Riesenhuber & Poggio 1999, Zhou et al. 2000). Even when using this experimental design, though, the persistent challenge of handling various more intricate visual problems, such as lightness and size illusions, motion aftereffects, or color constancy, indicated that a complete treatment of visual processing would require a more complex framework (Brainard & Freeman 1997, Chaudhuri 1990, Feldman 1997, Gilchrist et al. 1999, Palmer & Rock 1994, Wagemans et al. 2012). We posit that, with the expansion of the field of vision research to mid- and high-level vision, including Gestalt structures, surface perception, face recognition, and scene interpretation, and the realization of the important effects of context, biases, and the actual task for which vision is momentarily used, the contemporary definition of the problem of vision has changed from a problem of signal evaluation to a problem of goal-oriented interpretation (Froudarakis et al. 2019, Gilbert & Li 2013, Hayhoe 2017, Roelfsema & de Lange 2016, Yuille & Kersten 2006).

One of the most profound consequences among the many generated by this shift in approach concerns the status of internal knowledge and of the process of acquiring this knowledge, namely learning. Learning is commonly defined as a long-term improvement in performance due to

training or exposure (Fahle & Poggio 2002, Gallistel 1990, Gibson 1969, Sagi & Tanne 1994), and it has traditionally been treated as a fringe topic in vision, without which visual perception can be perfectly well understood (DeValois & DeValois 1990, Frisby & Stone 2010, Graham 1989). However, given that internally stored knowledge can radically and intricately influence the interpretation (i.e., the effect) of a piece of incoming sensory information even at the earliest visual areas (Briggs et al. 2013, Grosof et al. 1993, Kok & de Lange 2014, Kok et al. 2012, Murray et al. 2002, Smith & Muckli 2010, van Bergen et al. 2015), and that this knowledge perpetually changes due to recently obtained incoming information (Hua et al. 2010, Schoups et al. 2001), internal knowledge and learning, which define the emergence and change of this knowledge, become inseparable aspects of the visual process. The realization of this synergetic link gives a new importance to the topic of learning in vision.

In this review, we codify three types of learning, perceptual learning (PL), statistical learning (SL), and rule/abstract learning (RAL), and review their role in and the basis of their original separation in studies of visual perception. We argue that they are not separate learning types, but instead are extreme versions of the same process, and that learning in vision should be conceptualized in a framework that allows all three types of learning to occur simultaneously to capture human visual perception at the level of complexity where new insights could be gained about it. We propose that an extended version of SL firmly embedded in a probabilistic computational approach is the right conceptualization for such a framework. Finally, looking at the present literature through the lens of this framework, we identify what critical questions this extended SL framework should tackle next to unfold the full complexity of the interplay between visual perception and learning.

2. TYPES OF LEARNING IN VISION

2.1. Perceptual Learning

Learning in vision has been studied at different levels and can be roughly grouped into three areas with increasing levels of abstractness of the represented information: PL, SL, and RAL (**Figure 1**). PL is considered to be the most elementary type of learning, in which observers' performance improves in simple sensory tasks after extensive practice (Fahle & Poggio 2002) (**Figure 1*a***). Several comprehensive reviews are available on PL in vision (Dosher & Lu 2017, Maniglia & Seitz 2018, Sagi 2011, Watanabe & Sasaki 2015); thus, we mention only a few

characteristics relevant to our argument. Visual PL tasks typically include basic dimensions such as visual contrast (Adini et al. 2004, Dorais & Sagi 1997, Lengyel & Fiser 2019, Yu et al. 2004) and motion detection (Ball & Sekuler 1987), orientation (Fiorentini & Berardi 1980, Lengyel & Fiser 2019), or texture discrimination (Ahissar & Hochstein 1997). The effect of PL emerges after practice for 5–14 days of repetitive exposure over a couple of hours (Jeter et al. 2010), and it is quantified by improvements in a detection or discrimination threshold indicating a change in sensitivity (Fahle & Poggio 2002) (**Figure 1*b***). PL is considered to be a low-level phenomenon based on the specificity of learning, as any gain in performance that the learner demonstrates after practice would diminish as soon as the original experimental conditions are changed (Fahle & Morgan 1996). Several such conditions have been identified, such as presentation of the stimulus at a different location (Fahle & Morgan 1996, Schoups et al. 1995), orientation (Crist et al. 1997), or spatial frequency (Fiorentini & Berardi 1980), and performance decrement due to presentation of the monocular stimuli to a different eye has been fundamental in solidifying PL as a low-level process, since the integration of monocular representations happens in V1 (Schoups et al. 1995).

<COMP: PLEASE INSERT FIGURE 1 HERE>

Figure 1 The typical paradigms and corresponding results of perceptual learning (PL), statistical learning (SL), and rule/abstract learning (RAL). (*a*) PL using a classic orientation discrimination task with oriented grating stimuli. (*b*) Observers' discrimination thresholds improve over the course of the training session (*blue line*). Depending on the task and the stimuli, the discrimination thresholds at the beginning of a generalization task may remain high, indicating specificity (*red line*), or start at a low value, demonstrating transfer (*green line*). (*c*) Classical spatial visual SL with passive exposure to a stream of multi-element training scenes (shown on a gray background) generated from the inventory of chunks (*inset*) and subsequent 2-alternative-forced-choice familiarity test establishing the amount of learning. (*d*) Higher-than-chance (0.5) performance on the familiarity test indicates generalization through the learned chunks from the training to the test scenes. (NS: nonsignificant result) (*e*) Examples of RAL tasks in which, similar to classical temporal SL tasks, three groups of observers watch a temporal sequence of shape images but with different latent structures (triplets, maps, and networks, respectively). After being trained on one structure, the observers are exposed to examples of all three types of structure composed of elements that they have never seen before. (*f*) Observers learn the new task faster and better when the type of structure is similar to the one that they were exposed to during training.

2.2. Rule/Abstract Learning

RAL resides at the opposite end of the abstraction scale from PL (**Figure 1*e–f***). For a long time, it was considered to be outside of the scope of vision, as it deals with complex cognitive concepts based on abstract symbolic knowledge, including prototypes and higher-level features, and focuses on specific tasks, such as grammar learning (Dehaene et al. 2015, Fitch & Friederici 2012, Harlow 1949; Kemp & Tenenbaum 2008, Rabagliati et al. 2019, Saffran et al. 2007). Such learning is commonly referred to as rule learning owing to the fact that many complex concept-based structural descriptions are called rules, especially in research on language acquisition (Gómez & Gerken 1999, Marcus et al. 1999, Peña et al. 2002). However, a rule can be as simple as noting that a specific high tone is very likely to be accompanied by another specific low tone, which amounts to a high co-occurrence or transitional probability. Importantly, only the subset of learning such links that operates in dimensions clearly distanced from the observable sensory input would customarily qualify as abstract learning. Thus, while, for historical reasons, we refer to this type of learning as RAL, we emphasize that "rules" and "rule learning" are slightly confusing notational terms originating from research on language and logic, and that RAL learning in sensory research is better characterized by the level of abstraction and the degree of generalizability (Austerweil et al. 2019, Dehaene et al. 2015). Moreover, importing the term "rule learning" from formal grammar learning has led to also importing the problematic definition of abstract learning from the domain of language acquisition, which might not be appropriate for defining the same problem in vision. Therefore, we do not include these approaches anchored in the domain of language learning in the present overview beyond noting that many extensive reviews cover the topic (Aslin & Newport 2012, Dehaene et al. 2015, Fitch & Friederici 2012, Gómez & Gerken 2000)

Given the number of studies dealing with abstract learning in general, RAL has been linked to vision by a surprisingly small number of early behavioral work. An indication of the influence of the grammar learning approach is that the majority of these efforts were based on the paradigm used in research on infant grammar learning. These studies were typically based on the task of learning a repeating pattern of elements (e.g., AAB) in a sequential stream of spoken auditory input (Endress et al. 2007, Gerken 2006, Marcus et al. 1999, Peña et al. 2002). A common feature of these auditory studies was a strong separation of RAL from more basic learning types based on the fact that the test in the paradigm of these studies used completely new tokens not seen during the training session; thus, any learned information based directly on

the observed features of the training tokens could not serve as the basis of the acquired rule. The follow-up claim that, therefore, rule learning might be possible only for humans and only in the auditory domain because of our predisposition to learn languages (Marcus et al. 2004) has been proven incorrect through demonstrations of successful rule learning in the domain of vision in an identical experimental setup in both humans (Ferguson et al. 2018, Saffran et al. 2007) and rats (Murphy et al. 2008).

In other studies, RAL is attributed to vision mostly through visual cognition and scene interpretation without an organic link to visual perception (Gershman et al. 2016, Goodman et al. 2008, Lake et al. 2015, Mark et al. 2020, Overlan et al. 2017). Notable examples of recent RAL studies were based on quick learning of abstract regularities from exposure to a limited set of stimuli, and they tested generalization of this learning in adulthood and infancy (Buchsbaum et al. 2015, Ferguson et al. 2018, Garner et al. 2016, Mark et al. 2020, Overlan et al. 2017, Rabagliati et al. 2019, Werchan & Amso 2020) (**Figure 1*e***). In contrast to classical PL, where the dominant quantification of learning is the amount of improvement in performance, RAL is commonly evaluated in terms of the ability to generalize the acquired knowledge in a new context (Dehaene et al. 2015, Lake et al. 2017) (**Figure 1*f***). Generalization is typically measured either by the difference in performance at the beginning of the original and of the follow-up task (Wang et al. 2016) or by the speed of learning (Kattner et al. 2017). Crucially, while both the performance at the start of the test and the speed of learning can be observed in the shape of the measured learning curve, this curve is a simple one-dimensional aggregate indicator of a complex, multifactor learning process, which by itself is insufficient for uncovering the essential components of learning and their causal interactions (Heald et al. 2021). Indeed, RAL studies are labeled variously as investigations of "meta-learning," "learning-to-learn," "task-learning," "transfer-learning," or "structure learning," and this spectrum of labels indicates the wide range of causes that can determine the learning behavior under various conditions (Bavelier et al. 2012, Braun et al. 2010, Griffiths et al. 2019, Hupp & Sloutsky 2011, Kemp & Tenenbaum 2008, Kemp et al. 2010, Mark et al. 2020, Niv 2019, Schulz et al. 2020, Solway et al. 2014, Wang 2021, Woods and McDermott 2018).

Although RAL studies use a vast diversity of paradigms, types of inspected generalization, and applied tests, they can still be tabulated coarsely into two main groups by their focus. Studies in the first group explore the problem of how to represent knowledge in a structured manner for a

particular purpose. In this context, representing structure can mean separating out relevant from irrelevant features to form either simple or hierarchical categories of objects (Erdogan et al. 2015) or motion patterns (Bill et al. 2020), creating a map structure of a domain independent of tokens (Mark et al. 2020) or establishing the rules of a game regardless of nuisance parameters of the environment (Pouncy et al. 2021), but it can also refer to extracting aspects of a task in an experiment (Franklin & Frank 2018). Studies in the second group deal with the problem of how to use a structured representation for generalization purposes. These studies investigate how humans generalize through property induction (Kemp & Tenenbaum 2009), how they use learned reward functions for generalization during search tasks in spatially or conceptually correlated and graph-structured reward environments (Castañón et al. 2021; Wu et al. 2018, 2020), and how they can learn how to generalize (Austerweil et al. 2019).

A key characteristic of all of these studies is that, even if they are anchored in some sensory modality, their learning domain is structured and complex, they are defined by various abstract parameters and higher-level contexts, and they often include rewards or at least some partial feedback. In other words, the involved perceptual processes are always investigated in a context of a particular natural task, both in terms of the cover story and the experimental setup. Therefore, these studies convey information not about learning in a perceptual domain per se, but instead about learning in a perceptual domain given an abstract setting.

2.3. Statistical Learning

The more recently emerging third domain of learning, called SL, is situated between PL and RAL on the abstraction scale Fiser (2009) (**Figure 1*c–d***); however, its label is misleading on multiple counts. First, all learning is statistical, since all learning aims to extract structural information that, by definition, is manifested by various correlations in the input data and thus reflected by detectable statistics. Second, the term "statistical learning" has been used by mathematicians, statisticians, and computer scientists in a much wider computational context and different frameworks well before it emerged as a label with a restricted interpretation in the domain of language learning in infants (Hastie et al. 2013, Vapnik 1999, von Luxburg & Schölkopf 2011), and this reuse of the term has generated much confusion in the literature. Third, since its inception, research on SL has concentrated mainly on issues in the domain of language, and consequently, its profile has been heavily skewed toward abstract issues related to

the emergence of grammar in language acquisition (Aslin 2017, Saffran & Kirkham 2018, Saffran et al. 1996). Even though early papers transferring the basic methodology of SL into the domain of vision provided a setup suitable for breaking away from this restricted view of SL and integrating it with general statistics-based learning (Austerweil & Griffiths 2011, Fiser & Aslin 2005, Lee et al. 2021, Orbán et al. 2008, Yildirim & Jacobs 2013), to date, a substantial fraction of SL studies have followed the constrained path set by the early language-related work (Bettoni et al. 2021, Bulf et al. 2021, Frost et al. 2019, Schonberg et al. 2018, Siegelman et al. 2018).

The classical notion of SL is of a type of representational learning that is purely observational, without any explicit task or feedback, and that automatically and implicitly develops an internal structural representation of repeatedly appearing spatial and temporal patterns in the sensory input (Aslin 2017, Aslin & Newport 2012, Saffran & Kirkham 2018) (**Figure 1*c***). Early visual SL studies reported that both adults and infants become sensitive to joint, conditional probabilities and higher-order embedded spatial and temporal structures of previously unfamiliar inputs (Bulf et al. 2011; Fiser & Aslin 2002b, 2005; Ongchoco et al. 2016) (**Figure 1*d***). These early results were extended to several modalities (visual, auditory, and tactile) (Conway & Christiansen 2005, Glicksohn & Cohen 2013, Lengyel et al. 2019) and to different levels of stimulus complexities (Austerweil & Griffiths 2011, Orbán et al. 2008), and similar results were reported across several animal species (Avarguès-Weber et al. 2020, Rosa-Salva et al. 2018, Saffran et al. 2008, Santolin et al. 2016, Toro & Trobalón 2005). It has been firmly established that SL is automatic; its effect persists for a long time (Kim et al. 2009); sleep does not improve it (Nemeth et al. 2010); and, while attention can affect SL (Turk-Browne et al. 2005), it is not a prerequisite for successful learning (Musz et al. 2015).

At the computational level, the loose definition of SL and the preponderance of results across domains and conditions obtained with its methodology have led to several unresolved consequences. First, a lively debate has emerged over whether SL is a domain-specific or a domain-general process that might serve as the fundamental learning method for acquiring internal representations of the environment (Aslin 2017; Frost et al. 2015; Lengyel et al. 2019; Turk-Browne et al. 2005). Second, SL has been related to or equated with various other types of learning schemes, such as implicit learning (Perruchet & Pacton 2006); chunking

(Mareschal & French 2017, Perruchet 2019); probabilistic learning (Austerweil & Griffiths 2011, Orbán et al. 2008); and, through mixing Marr's levels of analysis, distributed learning and neural networks (Plaut & Vande Velde 2017, Schapiro et al. 2017). Third, as mentioned in Section 2.2, SL has been strongly separated from abstract rule learning and concept learning (Marcus et al. 1999, Peña et al. 2002). In addition, communities working with the separate types of PL, SL, and RAL have established their own stimuli, methodology, and measurement of learning, which prevents easy comparison and clarification of misunderstandings across these three subfields of research on learning. As a result, while each of these fields (especially the study of SL) has witnessed a spectacular increase in the number of publications and the variety of approaches over the years, there has been much less impressive progress in the conceptual clarification of how these different types of learning relate to each other and fit into perceptual processes.

3. THE NECESSITY OF INTEGRATING THE THREE TYPES OF LEARNING

In a recent paper, we investigated one aspect of integrating the three types of learning in vision by scrutinizing the relationship between PL and SL (Fiser & Lengyel 2019). We demonstrated that, with the advent of new and more complex behavioral experimental designs, the results of PL experiments started to show many signs of higher-level learning, losing their distinctiveness compared to SL; vice versa, the results of SL studies displayed effects at lower-level attributes that do not fit in the original symbolic framework. We went on to propose that PL and SL should be viewed not as two separate types of learning, but rather as two extreme testing paradigms of the same complex learning mechanism, where the PL paradigm lack more complex sensory structures and context, while the SL paradigm do not use low-level fine sensory features. Finally, we argued that, by treating PL and SL jointly in the framework of hierarchical Bayesian models (HBMs) and assuming a sampling-based approximative implementation of this framework in the brain, one could not only parsimoniously address outstanding puzzles in both fields but also provide several testable predictions about human learning at the theoretical as well as at the implementational level. Based on this conceptual setup, in this section, we investigate whether RAL should be integrated into the same framework and, if so, how.

3.1. Fundamental Characteristics of Statistical Learning

To evaluate the feasibility of integrating SL and RAL, we first assess the similarities and differences between the two types of learning. There are three essential features of the traditional SL paradigm: compositionality, implicitness, and abstraction. Compositionality means that the paradigm uses a limited number of observable tokens in a given sensory domain (e.g., $N < 20$ distinctive individual shapes in vision) and forms an inventory from these tokens in either the temporal or the spatial domain. In the temporal domain, a member of the inventory is a pair or triplet of shapes with elements that always appear consecutively in a fixed order, while in the spatial domain, the shapes of the pair or triplet always appear in a fixed spatial arrangement together. During a temporal practice session, such inventory elements are chained into a long stream without obvious separations between the inventory elements (Fiser & Aslin 2002a, Kirkham et al. 2002), while in the spatial practice, a few inventory elements are shown together in each scene without any segmentation cues separating the shapes of two inventory elements from each other (Fiser & Aslin 2001). This method of creating the sensory stimuli ensures that the identity of the inventory members (e.g., shape pairs) is never revealed while the individual elements (shapes) are all clearly visible. This way, while the set of stimuli has a well-defined underlying structure based on the inventory members, the observer sees only aggregate compositional scenes and never the underlying components alone. Moreover, due to careful counterbalancing in the experimental design, neither the features of the individual elements nor any other statistics of the scenes and streams (e.g., mean appearance frequency, position in the scene, or the identity of low-level features) can provide any relevant information about the underlying structure of the stimulus space.

The second feature, implicitness, means that, during the exposure period, the observers have no task to perform; they simply passively (but attentively) perceive the stream of stimuli containing the inventory elements from a few dozen to a hundred times. While this implicitness of the learning task should not be confused with the implicitness of the resulting knowledge, a large number of tests have confirmed that the overwhelming majority of existing SL studies produce explicit knowledge only in less than 5% of the observers (Bertels et al. 2012, Kim et al. 2009, Lengyel et al. 2019). This feature of the paradigm makes it suitable for investigating both issues of the effect of task implicitness or explicitness and the transition of knowledge from an implicit to an explicit state.

The third feature, abstractness, is related to the subsequent test following the exposure phase,

in which observers' familiarity with (Fiser & Aslin 2001) or reaction speed in response to (Barakat et al. 2013, Turk-Browne et al. 2005) true inventory elements versus random pair or triplet composition of shapes is assessed. Any difference in performance between true and random structures is taken as evidence that the observers learned to perform automatic segmentation of the input into sensible chunks during the exposure. In other words, they became more sensitive to the true inventory structure, i.e., they learned the members of the underlying inventory. Importantly, this is not an old–new test, since the measured sensitivity results from components not being seen alone during the exposure before the test; thus, the existence of these components has to be inferred from the composite scenes. In the simplest cases, SL tests can be passed by applying some straightforward counting strategies on elements or element pairs; experiments with more intricate designs showed that humans performed well in tests where such strategies were fruitless and true abstraction of the inventory members was required for success (Orbán et al. 2008). Nevertheless, with some notable exceptions (Fiser & Aslin 2005, Lee et al. 2021), the vast majority of the studies using this paradigm to date demonstrated learning of only the simplest statistics. These statistics comprise occurrence frequencies and joint as well as conditional (in the temporal domain, transitional) probabilities between two or three observable elements forming an inventory member (Bettoni et al. 2021, Bulf et al. 2011, Endress & Johnson 2021, Siegelman et al. 2019). Just as the ways in which humans learn higher-level statistics have not been fully explored, there are only a few studies focusing on other aspects of abstractness, i.e., the effect of multiple dimensions, context, and task dependency on SL (Luo & Zhao 2018, Otsuka & Saiki 2016, Turk-Browne et al. 2008, Zhao et al. 2011).

3.2. Fundamental Characteristics of Rule/Abstract Learning

As detailed above, RAL is a much less homogeneous domain than SL, but there are three essential features of RAL paradigms. These are the separation of learning from the observed stimulus, the explicitness of the observer's task, and the scope of higher-level abstraction. Separation from the observed stimulus means that the majority of abstract learning operates on concepts as inputs that do not have a direct equivalent in the sensory domains. For example, in the standard AAB rule-learning paradigm, even though observed tokens convey the relevant structure, the structure itself is defined along the abstract dimension of identity (same, same, different), and it can remain unchanged even if all observed tokens in the scene are replaced

(Saffran et al. 2007). This level of abstraction in RAL can be variable, however. For example, if the AAB structure is defined along a simpler dimension, such as size (i.e., small, small, large), several observed measures of the input, such as positional distribution of light energy in a given spatial frequency band, can strongly correlate with the abstract structure and thus can be used as a proxy for RAL (MacKenzie & Fiser 2010). In contrast, if the three items appear in different positions across the scenes, or if real-life sizes are considered instead of pictorial sizes, no low-level proxies can help, and learning must be performed at a truly higher level of abstraction.

The second feature, task explicitness, is just a typical feature, rather than an exclusive norm in RAL. For example, infant studies, by necessity, cannot use explicit tasks; thus, the AAB type and other infant studies rely on implicit tasks and, therefore, they are exceptions regarding this feature (Ferguson et al. 2018, Garner et al. 2016, Overlan et al. 2017, Rabagliati et al. 2019, Schonberg et al. 2018, Werchan & Amso 2020). Nevertheless, the majority of RAL studies with children and adults use verbal descriptions and specific cognitive tasks (e.g., categorization) to constrain the observers' learning processes (Franklin & Frank 2020, Rabi & Minda 2014, Yang et al. 2021, Yildirim & Jacobs 2013).

The third feature, the scope of higher-level abstraction, is the defining feature of RAL. Truly high-level abstraction is traditionally interpreted as mental processes related to ideas, i.e., abstract concepts that have no physical forms and can be handled by linguistic or amodal representations (Chomsky 1956, Dehaene et al. 2015, Lake et al. 2015, Pinker & Jackendoff 2005). Typical examples of such concepts include freedom, quality, humor, tradition, morals, mathematics, success, and learning, to name a few. In contrast, higher-level abstraction attributed to RAL is typically related to more concrete concepts, such as table, letter, or animal, that are more closely grounded in perception and action; possess some links to specific sensory features despite the large variability in appearance; and are frequently conceptualized as categories or object classes (Rosch 1973, 1975). Importantly, while the natural domain of RAL is higher- rather than truly high-level abstraction, this higher-level abstraction in RAL is more broadly defined than a generic taxonomic tree of category structures. As detailed in Section 2.2, this abstraction involves learning and utilizing relations, fragments, whole contexts, and task similarities and other structural information (Bavelier et al. 2012, Kattner et al. 2017, Kiefer & Harpaintner 2020).

3.3. The Case for Integrating Statistical Learning and Rule/Abstract Learning

Given these disparate characteristics of SL and RAL, why should they be integrated, and how does this integration relate to understanding visual perception? Early pioneers in vision research fully acknowledged the complexity of vision but followed the strategy of exploring the visual process only in an extremely restricted context and in isolation from higher-level processes that leads to object recognition, scene interpretation, or execution of various complex tasks based on vision. This approach was based on the rationale that a generic description of early vision might provide sufficient support for those higher-level cognitive investigations (Marr 1982). However, this strategy did not prevail, as it turned out that the context in which each isolated piece of low-level sensory information could appear strongly modulated the meaning and significance of the given information beyond any easy description (Cicchini et al. 2021, Cloherty et al. 2016, Grosof et al. 1993, Kok & de Lange 2014, Kok et al. 2012, Murray et al. 2002, Smith & Muckli 2010, van Bergen et al. 2015). Moreover, these contextual and task-related modulations could leave lasting effects in subsequent visual processes (Ahissar & Hochstein 1997, Ishikawa & Mogi 2011, Pomerantz et al. 1977). While these more complex results of vision solidified the notion that some learning must be handled together with visual perception in an integrated manner, it left open the question of which types of learning those are.

While superficial comparison of the essential features of SL and RAL gives the impression that it is correct to separate them in studies of vision, a more careful examination supports a different conclusion. Specifically, enforcement of compositionality at the level of tokens in SL paradigms suggests that extracting a more global structure of the input beyond the sea of pixel-wise correlations is an essential feature of SL. This means that SL performs abstraction away from the visually observed low-level attributes the same way that RAL does. While RAL focuses on higher levels of abstraction compared to SL, evidence suggests that abstraction can be defined on a continuous spectrum due to its being embedded in different perceptual processes to different degrees; this continuity removes the strict separation between RAL and SL (Fiser & Lengyel 2019, Kiefer & Harpaintner 2020). This view gained further support from a meta-analysis mapping existing PL, SL, and RAL studies along two axes: the complexity of the stimulus and the specificity of the task used in the study (**Figure 2*a***). As pointed out above, classical PL and SL studies used complementary types of stimuli (simple versus more complex displays) and opposite types of tasks (highly specific versus nonspecific), and both of these differences

reinforced the idea that PL has higher specificity, whereas SL has stronger generalization (Fiser & Lengyel 2019). With the introduction of newer paradigms in both domains that deviated from the classical tasks and stimuli, this distinction in generalization diminished, which led to the vanishing of the separation between PL and SL (Fiser & Lengyel 2019). While the stimulus complexity of SL and RAL studies are more comparable, RAL studies have a larger spectrum of task definitions compared to SL studies, ranging between implicit and explicit setups and focusing more on context and task effects. This detaches abstraction from generalization, and blurs the difference in the level of generalization between SL and RAL. In fact, some RAL studies using very specific task settings find minimal generalization (e.g., Kattner et al. 2016). However, forcing SL and RAL studies into implicit versus explicit task setups is not necessary, as investigation of both types of learning can easily begin with an implicit setup and continue toward explicit settings, which progression is a typical condition in natural tasks. This means that SL and RAL are neighbors in terms of the complexity of visual stimuli and can be defined very similarly in terms of task specificity.

<COMP: PLEASE INSERT FIGURE 2 HERE>

Figure 2 (*a*, *b*) The three learning types, indicated with colored labels and pictograms. (*a*) The relationships among the three learning types in terms of stimulus complexity and task specificity. Studies used in perceptual learning (PL) (*pink area*), statistical learning (SL) (*blue area*), and rule/abstract learning (RAL) (*green area*) are mapped onto the dimensions of stimulus complexity (*x* axis) and task specificity (*y* axis). The gray curved arrow indicates the direction of increased generalization reported in studies. The orange area depicts the range of natural tasks. (*b*) The relationships among the three learning types in terms of neural correlates and related brain areas. Reports on neural correlates of PL (*red*), SL (*blue*), and RAL (*green*) are ordered according to the complexity of the reported neural correlates modulated by learning (*x* axis) and approximate position of the investigated brain area within the cortical hierarchy (*y* axis) and colored in red, blue, or green according to the type of learning predominantly influencing the area. Colored dashed ellipses indicate typical combinations of neural correlates and involved areas of the three learning types. Bracketed numbers indicate the studies listed in the tables on the right. (*c*) Demonstration of the extended SL framework. The hierarchical Bayesian model (HBM) (*left*) provides a general schema for the computational framework unifying the three types of learning. An example of the extended SL paradigm based on the HBM involving PL (*bottom row*, *pink background*), SL (*middle row*, *blue background*), and RAL (*top row*, *green background*) is also shown along with the corresponding generative model formulated in the HBM framework (*framed*, *right*). In this example, RAL handles two task conditions changing randomly, trial by trial, between spatial frequency and orientation discrimination. SL handles the fact that, in each task, the reference value is selected, not randomly, but according to the order

defined by sequentially chosen reference pairs from an inventory. PL is responsible for the improvement of the fine discrimination between two sequentially provided Gabor stimuli.

Based on the above evidence about the different types of learning and the emerging link between learning and vision, the fundamental tenet of this review is that PL, SL, and RAL not only can be but also should be studied in a unified computational framework. If learning in perception is perpetual and occurs at each level of the representational hierarchy, and if the main goal of understanding learning is to clarify how it interacts with perception, particularly with vision, for higher efficiency, then a framework is necessary that can accommodate the various aspects of learning at all levels. In addition, this framework should naturally integrate with the interpretative definition of visual perception. Such a framework will have to embrace behavioral changes of experience-based learning, from the simplest sensitivity change in detecting contrast variations, to developing new general concepts such as firmness based on the visual appearance of a surface, to forming abstract categories based on other mental concepts such as task structures and contexts.

4. NEURAL CORRELATES OF LEARNING IN VISION

While reviewing the large spectrum of neural correlates in connection with each of the three learning types is beyond the scope of this review, several such reviews exist for PL (LeMessurier & Feldman 2018, Maniglia & Seitz 2018), SL (Batterink et al. 2019, Kourtzi & Welchman 2019), and RAL (Dehaene et al. 2015, Tervo et al. 2016). In this section, we point out the clear structural tendencies that can be observed when these neural signatures of the classical versions of the three learning types are tabulated along the axes of the neural correlates and the brain areas involved in learning (**Figure 2*b***). Classical PL studies typically found effects of learning in terms of simple single-unit measures such as changes in the mean, variability (tuning curves), or covariability of single-cell responses in lower visual areas. Meanwhile, early SL studies focusing on neural correlates reported involvements of only a few higher-order visual areas but many more areas not directly related to visual processing. SL studies also typically measured population activities and the strength of links between areas rather than firing rate means and variance. The neurophysiological results of RAL studies amplified these tendencies of SL studies by implicating a large number of high cognitive areas in the prefrontal and

orbitofrontal areas and finding effects of learning in terms of complex changes in functional connectivity across areas, rather than shifts in individual cell responses.

It is important to realize that the observed structural tendencies in **Figure 2*b*** are strongly related to the fact, demonstrated in **Figure 2*a***, that classical studies of PL, SL, and RAL use characteristically different types of stimuli and experimental tasks. Since learning is functionally defined by the tasks and the stimulus, the articulated separation of low- versus high-level learning types and the corresponding clustering of neural correlates should diminish as studies of the different learning types use more similar setups. Indeed, evidence collected with novel task paradigms and stimuli shows involvement of not only low-level but also higher-level cortical areas in PL (Jing et al. 2021, Law & Gold 2008, Li 2016), as well as neural signatures that are not related to receptive field changes but rather to altered population coding even in the primary visual areas (Ghose et al. 2002). This reinforces the idea that the strict separation between cortical areas involved in PL and SL is not warranted (Fiser & Lengyel 2019).

Research in the domain of spatial navigation has similarly diminished the separation between SL and RAL (Garvert et al. 2017, Hafting et al. 2005, Retailleau & Morris 2018). Based on the concept of cognitive maps (O'Keefe & Nadel 1978, Tolman 1948) and the discovery of place and grid cells (Hafting et al. 2005, O'Keefe & Dostrovsky 1971), a general framework of structural map-like internal representations in the brain emerged to explain 2D navigation (Nadasdy et al. 2017, O'Keefe & Nadel 1978). Learning representations for 2D navigation can be viewed as a process of learning more complex versions of the spatial structures handled by SL. This framework has been generalized from spatial to other map-like structures and further to any structural nonspatial internal representation based on evidence that neural activity in the hippocampal formation and that in the prefrontal cortex exhibit similar patterns (Baram et al. 2021). These representations were suggested to enable flexible human behavior by facilitating abilities of inference and abstraction on the type of structural knowledge that RAL generates (Mark et al. 2020).

5 THE COMPUTATIONAL FRAMEWORK OF LEARNING IN VISION

5.1. Implications of Previous Perceptual Learning, Statistical Learning, and Rule/Abstract Learning Models for a Common Framework

As computational models of PL have been reviewed extensively elsewhere (Dosher & Lu 2017,

Fiser & Lengyel 2019), in this section, we focus on the requirements of integrating SL and RAL into a single computational framework. Inspired by the agenda set in the study of language learning, early models of visual SL were framed in the context of learning pairs or triplets of elements in a stream of sequentially presented single items with the central question of whether learning transitional probabilities or identifying group of elements as chunks is the correct underlying computational model (Perruchet 2019, Perruchet & Pacton 2006). While this debate has been resolved by evidence that chunk learning is a better conceptualization (Glicksohn & Cohen 2011, Orbán et al. 2008, Perruchet 2019), different definitions of chunk and versions of the underlying computational models have emerged in the literature. These proposals range from models combining cooperation and competition principles with observed psychological process, such as compositionality of the representation (Perruchet & Vinter 1998); to connectionist implementation of autoassociative networks (French et al. 2011), i.e., neural network implementations with higher fidelity of cortical structures (Schapiro et al. 2017); to probabilistic models (Austerweil & Griffiths 2011,Lee et al. 2021, Orbán et al. 2008).**[**AU: Edits correct?**]** For example, recent improvements of the connectionist approach using deep neural networks (DNNs) showed remarkable results in solving difficult visual problems by distributed hierarchical supervised learning (Eickenberg et al. 2017, Kriegeskorte 2015, Kubilius et al. 2016, Wenliang & Seitz 2018, Yamins et al. 2014). However, there seems to be a considerable consensus that, despite this success, DNNs do not capture the essential features of human visual learning (Geirhos et al. 2018, Kietzmann et al. 2019, Lake & Baroni 2018, Lake et al. 2017, Srivastava et al. 2019, Ullman et al. 2016). In contrast, it has been found that, while simple visual SL results can be replicated even by frequency- or co-occurrence-counting naive models, only probabilistic chunk learning models can capture humans' SL results of learning more challenging embedded structures of the visual input (Orbán et al. 2008).

As discussed in Section 2.2, the study of RAL is not linked strongly to vision, as it focuses on the general principles of human learning, and therefore, it investigates the effects of tasks, contexts, and overall distal structures generating the sensory input, with less emphasis on the visual signal per se. Consequently, models of RAL utilize a wide range of current techniques of machine learning that are not reviewed in this article for the sake of brevity (Alhama & Zuidema

2019, Altmann 2002, Lake et al. 2017). Due to the large, complex, and structured problem spaces and the strong emphasis on flexible generalization as the main goal, the overwhelming majority of RAL models utilize a probabilistic framework (Acuña & Schrater 2010, Austerweil et al. 2019, Bavelier et al. 2012, Collins & Frank 2013, Eckstein & Collins 2020, Garvert et al. 2017, Lake et al. 2017, Mark et al. 2020, Tomov et al. 2021). These models, particularly HBMs, have been argued to be the most suitable for handling the challenges of abstract learning problems (Lake et al. 2015, 2017; Tenenbaum et al. 2011). The HBM has also been the framework of choice for integrating PL and SL to formalize the interplay between those two types of learning (Fiser & Lengyel 2019).

5.2. Integrating Perceptual Learning, Statistical Learning, and Rule/Abstract Learning in a Hierarchical Bayesian Model

The converging views in the studies of SL and RAL on using the probabilistic framework are in consonance with the recent trend of treating complex visual perception with probabilistic models (Knill & Pouget 2004, Yuille & Kersten 2006). These models represent both sensory information and uncertainty about that information, and thus, they can effectively handle the ambiguity and context dependence of the stimulus (Eckstein 2017, Hayhoe 2017, Murray 2021, Yuille & Kersten 2006). A recent paper argued that, given the scope of generalizability demonstrated by humans in visually guided natural tasks, uncertainty must be represented even at the earliest levels of the hierarchical sensory processing, and that evaluating the incoming information requires a joint inference across all of these levels (Koblinger et al. 2021). However, if this is the case, then learning must follow the same strategy by jointly adjusting the internal knowledge through coordinating across all three types of learning; otherwise, it cannot support vision effectively (Fiser et al. 2010). We posit that, for integrating learning and visual perception with the goal of better understanding complex vision, a model of learning is required that comprises all of the types of learning in a framework where they can seamlessly interact with each other and with the perceptual process.

We have presented a proof-of-concept example to demonstrate that this requirement is well satisfied by an HBM model in the case of integrating PL and SL (Fiser & Lengyel 2019). The SL part of this example was depicted by the simplest pair-based contextual structure, but HBM allows learning and using for inference a much richer structural representation (Fazeli et al.

2019; Lake et al. 2015, 2017). Computational studies have demonstrated that the higher levels of this structured representation can evolve to various representational arrangement classes ranging from simple linear 1D chains to more complicated 2D grids, trees, rings, and cliques (Kemp & Tenenbaum 2008). In principle, this representation can successfully capture not only the type of rules used in empirical rule-learning vision experiments to date, but also highly specific real structures or abstract regularities or laws based on spatial arrangements, lighting, or size and depth variations (French & DeAngelis 2020, Lee et al. 2021, Murray 2021, Orbán et al. 2008). Thus, RAL can be integrated into the proposed PL–SL framework simply by extending upward the hierarchy of the upper structure, which can store highly ordered internal knowledge traditionally not included in PL and SL studies (**Figure 2*c***). Such an integrated learning model can naturally discover and generalize across structural regularities at any level, from sensory variations through object and sequence identities to context, task, and value similarities that inescapably emerge in any complex visual situation.

5.3. The Blessing of Approximation-by-Sampling for Implementing a Probabilistic Framework

Interpreting learning through a wide variety of the measurements of neural responses has been one of the fundamental difficulties for establishing the link between vision and learning (LeMessurier & Feldman 2018). The proposal that learning in vision should be formalized by HBMs introduces abstract probability distributions as the fundamental representational concept for perception, instead of the traditionally used firing rates of individual cells, and this seems to make the challenge of interpretations even harder (Pouget et al. 2013). However, rather than complicating things, this shift in the proposed computational representations gives an opportunity for a reinterpretation and unification of earlier findings in the literature. The probability distributions used by the HBM framework are computationally intractable for real-world problems, and therefore, they require feasible approximative algorithms at the level of neural computing when implemented in the brain (Fiser et al. 2010, Pouget et al. 2013). Consequently, a principled mapping between the abstract probabilistic computations for learning (and visual perception) and the measurable neural characteristics of the brain's activity associated with learning needs to be established to link the framework to physiological quantities. Different frameworks including probabilistic population codes (Ma et al. 2006) and sampling-based methods (Fiser et al. 2010, Hoyer & Hyvarinen 2003, Lee & Mumford 2003)

have been proposed as biologically feasible approximations of probabilistic inference in the brain. Among these are sampling-based approximations, which have been argued to capture the available neural evidence not only for perception but also for learning (Fiser et al. 2010). In addition, recent papers have shown how sampling-based methods can derive a precise mapping between the abstract probabilistic computations and different traditional signatures of neural activity, including neural tuning curves, response means and variability, gains, correlations and population sparseness, response dynamics, and oscillations (Echeveste et al. 2020, Orbán et al. 2016). These results represent the first steps in developing appropriate mappings that can provide the necessary means to establish a systematic connection between the abstract formalism of perception and learning and the physiological characteristics of brain functioning at multiple levels of the neural hierarchy.

6. CHALLENGES OF INTEGRATING STATISTICAL LEARNING WITH VISION

As we claim in this review, an extended version of learning based on SL combined with PL and RAL in a probabilistic framework might be the right approach to integrate learning into vision; however, the currently available models and studies based on this principle represent an unquestionably modest initial step in this direction. In this final section of our review, we list five major challenges that SL faces in becoming a feasible and adequate model of visual learning and mention recent developments in each of these areas. The five challenges relate to suitable experimental stimuli, hierarchy of representation, link between SL and vision, formalization of learning in vision and finally, the neural correlates of visual learning.

The first challenge is equipping SL with more natural stimuli and context. The very first step is to eliminate the artificial separation between temporal (Turk-Browne et al. 2005) and spatial (Fiser & Aslin 2001) SL paradigms and to begin working with spatiotemporal sequences (Garber & Fiser 2021b). This would facilitate the abandonment of the use of simplistic and detached models based on transitional-probability/temporal-prediction versus batch-clustering/occurrence-frequencies and foster the emergence of models with adequate complexity to support visual perception. The second step is to use stimuli with natural dimensions (gray-scale images of real 3D shapes) rather than symbols (individual black 2D forms on a white background on a grid). Integrating PL and SL will necessarily speed up this transition, but a

formidable task will be to develop such stimuli without losing control over the increase in learning complexity due to the evoked number of dimensions. A corollary step in solving this challenge is using not only natural dimensions for the stimuli, but also natural structures of the visual input, such as occlusions and saccade-based changes.

The second challenge is integrating RAL and making it relevant to vision. As a first step, the underlying structural complexity of the setups used in SL paradigms should be extended to allow for the emergence of more abstract hierarchies of representations. These structures could include categories (Garber & Fiser 2021a) and contexts based both on input stimuli and on task specification (Collins & Frank 2013, Minda & Smith 2001, Werchan et al. 2015). The second step is to explore how the emergent abstractions by SL would tie into fundamental variables of visual perception. One possible example of this exploration is clarifying the links among the occlusion-based emergence of the concept of 3D depth, the emergence of absolute size based on depth, and the potential role of this abstraction in the emergence of size invariance and size constancy. Understanding such abstraction processes in learning would help us identify how abstract concepts contribute to visual perception.

The third, complementary, challenge is validating the representation resulted from SL as a true object-like representation. If SL is integrated with visual perception, then it should create representations that replicate known behavioral phenomena that emerge with true object representations. For example, one of the multiple well-documented attentional effects in vision, called object-based attention, has recently been reproduced with chunks learned in a standard SL paradigm (Lengyel et al. 2021). Another example is amodal object generalization in perception, for which a study has shown that learning of purely visual or purely haptic sensory structures by SL immediately generalizes over to the other modality (Lengyel et al. 2019). Similar studies are needed for several levels of visual processing ranging from Gestalt effects to illusions of various kinds. A related aspect of this challenge concerns the links between long-term semantic memory formed by SL and two other types of visual memory, working memory (WM) and episodic memory. Brady et al. (2009) provided initial reports of interaction between knowledge formed by SL and WM, but further clarifications are needed. While episodic memory in vision has been explored by measuring the capacity and specificity of memory for episodic information (Brady et al. 2008), the relationship between episodic memory and SL-based representations is a widely unexplored topic (Sherman & Turk-Browne 2020).

The fourth challenge is formally developing and validating the HBM computational framework that accommodates results of not only SL but also PL and RAL (**Figure 2*c***). As a first step, the joint modelling of PL and SL should be developed to explain the puzzling pattern of results in the literature obtained by nontraditional PL or SL paradigms. These models could use only one level of the upper hierarchy, representing the knowledge gained by SL, and show how the influence of this knowledge on lower-level stimuli provides the flexibility to capture human performance. This model structure would be sufficient to explore, for example, the roving results of PL, that is, the well-established phenomenon that, when a PL discrimination task is not learnable with the reference stimuli randomly interleaved during training, it is often learned when the reference stimuli are grouped in blocks or change according to a fixed sequence (Kuai et al. 2005, Zhang et al. 2008). The same model should be applicable to handle the enhanced generalization effects observed in double-training PL paradigms (Wang et al. 2014, Xiao et al. 2008), including category-induced biases in orientation perception (Tan et al. 2019) and even results showing imagination-based improvements in PL (Tartaglia et al. 2009). Moreover, the same model structure should be sufficient to handle results showing the perceptual biases of classical studies of SL (Barakat et al. 2013, Luo & Zhao 2018, Zhao et al. 2011).

The second step is integrating RAL into the framework by extending the number of levels in the upper part of the HBM and investigating more complex interactions between variables at the different levels of abstraction. These investigations can approach the issue from two directions. First, they can focus on clarifying whether there is evidence for hallmark characteristics of probabilistic representation and inference, such as explaining-away effects (Yuille & Kersten 2006), in hierarchical representations newly learned by SL. Second, they can explore whether the proposed framework could accommodate various specific types of metalearning effects in the literature, such as extensive play of video games causing improvements not only in spatial and temporal resolution and contrast sensitivity of vision, but also in visual short-term memory and information accumulation for decision making (Bavelier et al. 2012). The third formidable step in overcoming this challenge is incorporating the temporal aspect of sequentiality in learning into current models. Although various simple approximations of sequential learning, such as hidden Markov models, state space models, and partially observed Markov decision processes, already exist, a full framework of feasible approximation of sequential probabilistic learning in the brain is still in development (Heald et al. 2021, Radulescu et al. 2021).

The final challenge is identifying physiologically testable neural correlates of statistical learning in vision (https://www.kitp.ucsb.edu/activities/brainlearn23). Pinpointing the neural correlates of sensory learning has been a notoriously difficult undertaking, since effects of specific vision-related changes were mixed with those of general context- and task-related changes (Law & Gold 2008). With the advent of a more sophisticated probabilistic SL model of the learning process, the different effects might be parsed more successfully (Heald et al. 2021). While the first steps of deriving neurophysiologically meaningful predictions based on sophisticated approximative probabilistic models have been made (Echeveste et al. 2020, Orbán et al. 2016), the development of a much wider scope of metrics capturing complex neural phenomena is still a task for future research (Buzsáki & Draguhn 2004, Semedo et al. 2019).

SUMMARY POINTS

1. To understand complex vision, research on visual perception and learning needs to be integrated.
2. To exploit the potential insights that learning can offer to understand vision, the full spectrum of learning, currently represented by the separate fields of PL, SL, and RAL, must be considered as a whole.
3. Such a joint treatment is possible under an approximate probabilistic framework, since PL, SL, and RAL are not conceptually different learning types, but rather, different extreme versions of the same representational learning schema.
4. The unified framework can offer not only explanations for hitherto puzzling behavioral phenomena and new predictions, but also a tighter link between vision and the neural bases of learning.

FUTURE ISSUES

1. Integration of vision and learning requires the extension and synthesis of existing visual learning paradigms. This includes the use of more complex and more natural stimuli with a hierarchical structure, as well as consideration of the context of the visual input.
2. In addition, the relevance of the established link between vision and learning needs to be

confirmed by experiments linking the representations resulted from visual learning to true object representations.

3. The adequateness of the hierarchical probabilistic framework for visual learning needs to be clarified computationally with an emphasis on sequential learning and the amount and type of generalization that the framework can provide.
4. The biological plausibility of the proposed framework has to be established by providing a general neural coding approach that can capture, in a feasible manner, probabilistic computation in the brain and provide testable predictions for electrophysiological experiments.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Acuña DE, Schrater P. 2010. Structure learning in human sequential decision-making. *PLOS Comput. Biol.* 6(12):e1001003

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):401–6

Alhama RG, Zuidema W. 2019. A review of computational models of basic rule learning: the neural-symbolic debate and beyond. *Psych. Bull. Rev.* 26(4):1174–94

Altmann GTM. 2002. Learning and development in neural networks—the importance of prior experience. *Cognition* 85(2):B43–50

Aslin RN. 2017. Statistical learning: a powerful mechanism that operates by mere exposure. *Wiley Interdiscip. Rev. Cogn. Sci.* 8(1–2):e1373

Aslin RN, Newport EL. 2012. Statistical learning: from acquiring specific items to forming general rules. *Curr. Dir. Psychol. Sci.* 21(3):170–76

Austerweil JL, Griffiths TL. 2011. A rational model of the effects of distributional information on feature learning. *Cogn. Psychol.* 63(4):173–209

Austerweil JL, Sanborn S, Griffiths TL. 2019. Learning how to generalize. *Cogn. Sci.* 43(8):e12777

Avarguès-Weber A, Finke V, Nagy M, Szabó T, d'Amaro D, et al. 2020. Different mechanisms underlie implicit visual statistical learning in honey bees and humans. *PNAS* 117(41):25923–34

Ball K, Sekuler R. 1987. Direction-specific improvement in motion discrimination. *Vis. Res.* 27(6):953–65

Barakat BK, Seitz AR., Shams L. 2013. The effect of statistical learning on internal stimulus representations: Predictable items are enhanced even when not predicted. *Cognition* 129(2):205–11

Baram AB, Muller TH, Nili H, Garvert MM, Behrens TEJ. 2021. Entorhinal and ventromedial prefrontal cortices abstract and generalize the structure of reinforcement learning problems. *Neuron* 109(4):713–23.e7

Batterink LJ, Paller KA, Reber PJ. 2019. Understanding the neural bases of implicit and statistical learning. *Topics Cogn. Sci.* 11(3):482–503

Bavelier D, Green CS, Pouget A, Schrater P. 2012. Brain plasticity through the life span: learning to learn and action video games. *Annu. Rev. Neurosci.* 35:391–416

Bertels J, Franco A, Destrebecqz A. 2012. How implicit is visual statistical learning? *J. Exp. Psychol. Learn. Mem. Cogn.* 38(5):1425–31

Bettoni R, Bulf H, Brady S, Johnson SP. 2021. Infants' learning of non-adjacent regularities from visual sequences. *Infancy* 26(2):319–26

Bill J, Pailian H, Gershman SJ, Drugowitsch J. 2020. Hierarchical structure is employed by humans during visual motion perception. *PNAS* 117(39):24581–89

Brady TF, Konkle T, Alvarez GA. 2009. Compression in visual working memory: using statistical regularities to form more efficient memory representations. *J. Exp. Psychol. Gen.* 138(4):487–502

Brady TF, Konkle T, Alvarez GA, Oliva A. 2008. Visual long-term memory has a massive storage capacity for object details. *PNAS* 105(38):14325–29

Brainard DH, Freeman WT. 1997. Bayesian color constancy. *J. Opt. Soc. Am. A* 14(7):1393–411

Braun DA, Mehring C, Wolpert DM. 2010. Structure learning in action. *Behav. Brain Res.* 206(2):157–65

Briggs F, Mangun GR, Usrey WM. 2013. Attention enhances synaptic efficacy and the signal-to-noise ratio in neural circuits. *Nature* 499(7459):476–80

Buchsbaum D, Griffiths TL, Plunkett D, Gopnik A, Baldwin D. 2015. Inferring action structure and causal relationships in continuous sequences of human action. *Cogn. Psychol.* 76:30–77

Bulf H, Johnson SP, Valenza E. 2011. Visual statistical learning in the newborn infant. *Cognition* 121(1):127–32

Bulf H, Quadrelli E, Brady S, Nguyen B, Cassia VM, Johnson SP. 2021. Rule learning transfer across linguistic and visual modalities in 7-month-old infants. *Infancy* 26(3):442–54

Buzsáki G, Draguhn A. 2004. Neuronal oscillations in cortical networks. *Science* 304(5679):1926–29

Castañón SH, Cardoso-Leite P, Altarelli I, Green CS, Schrater P, Bavelier D. 2021. A mixture of generative models strategy helps humans generalize across tasks. bioRxiv 2021.02.16.431506. https://doi.org/10.1101/2021.02.16.431506

Chaudhuri A. 1990. Modulation of the motion aftereffect by selective attention. *Nature* 344(6261):60–62

Chomsky N. 1956. Three models for the description of language. *IEEE Trans. Inf. Theory* 2(3):113–24

Cicchini GM, Benedetto A, Burr DC. 2021. Perceptual history propagates down to early levels of sensory analysis. *Curr. Biol.* 31(6):1245–50.e2

Cloherty SL, Hughes NJ, Hietanen MA, Bhagavatula PS, Goodhill GJ, Ibbotson MR. 2016. Sensory experience modifies feature map relationships in visual cortex. *eLife* 5:e13911

Collins AGE, Frank MJ. 2013. Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychol. Rev.* 120(1):190–229

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

Crist RE, Kapadia MK, Westheimer G, Gilbert CD. 1997. Perceptual learning of spatial localization: specificity for orientation, position, and context. *J. Neurophysiol.* 78(6):2889–94

de Lange FP, de Lange FP, Heilbron M, Kok P. 2018. How do expectations shape perception? *Trends Cogn. Sci.* 22(9):P764–79

Dehaene S, Meyniel F, Wacongne C, Wang L, Pallier C. 2015. The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron* 88(1):2–19

DeValois RL, DeValois KK. 1990. *Spatial Vision*. Oxford, UK: Oxford Univ. Press

DiCarlo JJ, Zoccolan D, Rust NC. 2012. How does the brain solve visual object recognition? *Neuron* 73(3):415–34

Dorais A, Sagi D. 1997. Contrast masking effects change with practice. *Vis. Res.* 37(13):1725–33

Dosher B, Lu Z-L. 2017. Visual perceptual learning and models. *Annu. Rev. Vis. Sci.* 3:343–63

Echeveste R, Aitchison L, Hennequin G, Lengyel M. 2020. Cortical-like dynamics in recurrent circuits optimized for sampling-based probabilistic inference. *Nat. Neurosci.* 23(9):1138–49

Eckstein MK, Collins AGE. 2020. Computational evidence for hierarchically structured reinforcement learning in humans. *PNAS* 117(47):29381–89

Eckstein MP. 2017. Probabilistic computations for attention, eye movements, and search. *Annu.*

Rev. Vis. Sci. 3:319–42

Eickenberg M, Gramfort A, Varoquaux G, Thirion B. 2017. Seeing it all: Convolutional network layers map the function of the human visual system. *NeuroImage* 152:184–94

Endress AD, Dehaene-Lambertz G, Mehler J. 2007. Perceptual constraints and the learnability of simple grammars. *Cognition* 105(3):577–614

Endress AD, Johnson SP. 2021. When forgetting fosters learning: a neural network model for statistical learning. *Cognition* 213:104621

Erdogan G, Yildirim I, Jacobs RA. 2015. From sensory signals to modality-independent conceptual representations: a probabilistic language of thought approach. *PLOS Comput. Biol.* 11(11):e1004610

Fahle M, Morgan M. 1996. No transfer of perceptual learning between similar stimuli in the same retinal position. *Curr. Biol.* 6(3):292–97

Fahle M, Poggio TA. 2002. *Perceptual Learning*. Cambridge, MA: MIT Press

Fazeli N, Oller M, Wu J, Wu Z, Tenenbaum JB, Rodriguez A. 2019. See, feel, act: hierarchical learning for complex manipulation skills with multisensory fusion. *Sci. Robot.* 4(26):eaav3123

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

Ferguson B, Franconeri SL, Waxman SR. 2018. Very young infants learn abstract rules in the visual modality. *PLOS ONE* 13(1):e0190185

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

Fiser J. 2009. Perceptual learning and representational learning in humans and animals. *Learning & behavior 37 (2)*:141-153

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

Fiser J, Aslin RN. 2002a. Statistical learning of higher-order temporal structure from visual

shape sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 28(3):458–67

Fiser J, Aslin RN. 2002b. Statistical learning of new visual feature combinations by infants. *PNAS* 99(24):15822–26

Fiser J, Aslin RN. 2005. Encoding multielement scenes: statistical learning of visual feature hierarchies. *J. Exp. Psychol. Gen.* 134(4):521–37

Fiser J, Berkes P, Orbán G, Lengyel M. 2010. Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn. Sci.* 14(3):119–30

Fiser J, Lengyel G. 2019. A common probabilistic framework for perceptual and statistical learning. *Curr. Opin. Neurobiol.* 58:218–28

Fitch WT, Friederici AD. 2012. Artificial grammar learning meets formal language theory: an overview. *Philos. Trans. R. Soc. Lond. B* 367(1598):1933–55

Franklin NT, Frank MJ. 2018. Compositional clustering in task structure learning. *PLOS Comput. Biol.* 14(4):e1006116

Franklin NT, Frank MJ. 2020. Generalizing to generalize: Humans flexibly switch between compositional and conjunctive structures during reinforcement learning. *PLOS Comput. Biol.* 16(4):e1007720

French RL, DeAngelis GC. 2020. Multisensory neural processing: from cue integration to causal inference. *Curr. Opin. Physiol.* 16:8–13

French RM, Addyman C, Mareschal D. 2011. TRACX: a recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychol. Rev.* 118(4):614–36

Frisby JP, Stone JV. 2010. *Seeing: The Computational Approach to Biological Vision*. Cambridge, MA: MIT Press

Frost R, Armstrong BC, Christiansen MH. 2019. Statistical learning research: a critical review and possible new directions. *Psychol. Bull.* 145(12):1128–53

Frost R, Armstrong BC, Siegelman N, Christiansen MH. 2015. Domain generality versus modality specificity: the paradox of statistical learning. *Trends Cogn. Sci.* 19(3):117–25

Froudarakis E, Fahey PG, Reimer J, Smirnakis SM, Tehovnik EJ, Tolias AS. 2019. The visual cortex in context. *Annu. Rev. Vis. Sci*. 5:317–39

Gallistel CR. 1990. *The Organization of Learning*. Cambridge, MA: MIT Press

Garber D, Fiser J. 2021a. Pre-training leads to a structural novelty effect in spatial visual statistical learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, art. 43. N.p.: Cogn. Sci. Soc. https://escholarship.org/content/qt9qc0x5n1/qt9qc0x5n1.pdf?t=qwi3u0

Garber D, Fiser J. 2021b. Recovering spatial structure in spatio-temporal visual statistical learning. *J. Vis*. 21(9):2160

Garner KG, Lynch CR, Dux PE. 2016. Transfer of training benefits requires rules we cannot see (or hear). *J. Exp. Psychol. Hum. Percept. Perform.* 42(8):1148–57

Garvert MM, Dolan RJ, Behrens TE. 2017. A map of abstract relational knowledge in the human hippocampal-entorhinal cortex. *eLife* 6:e17086

Geirhos R, Medina Temme CR, Rauber J, Schutt HH, Bethge M, Wichmann FA. 2018. Generalisation in humans and deep neural networks. In *NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 7549–61. Red Hook, NY: Curran Assoc.

Gerken L. 2006. Decisions, decisions: infant language learning when multiple generalizations are possible. *Cognition* 98(3):B67–74

Gershman SJ, Tenenbaum JB, Jäkel F. 2016. Discovering hierarchical motion structure. *Vis. Res.* 126:232–41

Ghose GM, Yang T, Maunsell JHR. 2002. Physiological correlates of perceptual learning in monkey V1 and V2. *J. Neurophysiol.* 87(4):1867–88

Gibson EJ. 1969. *Principles of Perceptual Learning and Development*, Vol. 6. New York: Appleton-Century-Crofts

Gilbert CD, Li W. 2013. Top-down influences on visual processing. *Nat. Rev. Neurosci.* 14:350–63

Gilchrist A, Kossyfidis C, Bonato F, Agostini T, Cataliotti J, et al. 1999. An anchoring theory of lightness perception. *Psychol. Rev*. 106(4):795–834

Glicksohn A, Cohen A. 2011. The role of Gestalt grouping principles in visual statistical learning. *Attention Percept. Psychophys.* 73(3):708–13

Glicksohn A, Cohen A. 2013. The role of cross-modal associations in statistical learning. *Psychon. Bull. Rev.* 20(6):1161–69

Gómez RL, Gerken L. 1999. Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* 70(2):109–35

Gómez RL, Gerken L. 2000. Infant artificial language learning and language acquisition. *Trends Cogn. Sci.* 4(5):P178–86

Goodman ND, Tenenbaum JB, Feldman J, Griffiths TL. 2008. A rational analysis of rule-based concept learning. *Cogn. Sci.* 32(1):108–54

Graham N. 1989. *Visual Pattern Analyzers*. Oxford, UK: Oxford Univ. Press

Griffiths TL, Callaway F, Chang MB, Grant E, Krueger PM, Lieder F. 2019. Doing more with less: meta-reasoning and meta-learning in humans and machines. *Curr. Opin. Behav. Sci*. 29:24–30

Grosof DH, Shapley RM, Hawken MJ. 1993. Macaque V1 neurons can signal "illusory" contours. *Nature* 365(6446):550–52

Hafting T, Fyhn M, Molden S, Moser M-B, Moser EI. 2005. Microstructure of a spatial map in the entorhinal cortex. *Nature* 436 (7052):801–6

Harlow HF. 1949. The formation of learning sets. *Psychol. Rev.* 56(1):51–65

Hastie T, Tibshirani R, Friedman J. 2013. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Berlin: Springer

Hayhoe MM. 2017. Vision and action. *Annu. Rev. Vis. Sci*. 3:389–413

Heald JB, Lengyel M, Wolpert DM. 2021. Contextual inference underlies the learning of sensorimotor repertoires. *Nature* 600:489–93

Hoyer PO, Hyvarinen A. 2003. Interpreting neural response variability as Monte Carlo sampling of the posterior. In *Advances in Neural Information Processing Systems 15*, pp. 293–300. Cambridge, MA: MIT Press

Hua T, Bao P, Huang C-B, Wang Z, Xu J, et al. 2010. Perceptual learning improves contrast sensitivity of V1 neurons in cats. *Curr. Biol.* 20(10):887–94

Hupp JM, Sloutsky VM. 2011. Learning to learn: from within-modality to cross-modality transfer during infancy. *J. Exp. Child Psychol.* 110(3):408–21

Ishikawa T, Mogi K. 2011. Visual one-shot learning as an "anti-camouflage device": a novel morphing paradigm. *Cogn. Neurodyn.* 5(3):231–39

Jeter PE, Dosher BA, Liu S-H, Lu Z-L. 2010. Specificity of perceptual learning increases with increased training. *Vis. Res.* 50(19):1928–40

Jing R, Yang C, Huang X, Li W. 2021. Perceptual learning as a result of concerted changes in prefrontal and visual cortex. *Curr. Biol.* 31(20):P4521–33.E3

Kattner F, Cochrane A, Cox CR, Gorman TE, Green CS. 2017. Perceptual learning generalization from sequential perceptual training as a change in learning rate. *Curr. Biol.* 27(6):840–46

Kattner F, Cox CR, Green CS. 2016. Transfer in rule-based category learning depends on the training task. *PLOS ONE* 11(10):e0165260

Kemp C, Goodman ND, Tenenbaum JB. 2010. Learning to learn causal models. *Cogn. Sci.* 34(7):1185–243

Kemp C, Tenenbaum JB. 2008. The discovery of structural form. *PNAS* 105(31):10687–92

Kemp C, Tenenbaum JB. 2009. Structured statistical models of inductive reasoning. *Psychol. Rev.* 116(1):20–58

Kiefer M, Harpaintner M. 2020. Varieties of abstract concepts and their grounding in perception or action. *Open Psychol*. 2(1):119–37

Kietzmann TC, Spoerer CJ, Sörensen LKA, Cichy RM, Hauk O, Kriegeskorte N. 2019. Recurrence is required to capture the representational dynamics of the human visual system.

PNAS 116(43):21854–63

Kim R, Seitz A, Feenstra H, Shams L. 2009. Testing assumptions of statistical learning: Is it long-term and implicit? *Neurosci. Lett.* 461(2):145–49

Kirkham NZ, Slemmer JA, Johnson SP. 2002. Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* 83(2):B35–42

Knill DC, Pouget A. 2004. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27(12):712–19

Koblinger Á, Fiser J, Lengyel M. 2021. Representations of uncertainty: Where art thou? *Curr. Opin. Behav. Sci.* 38:150–62

Kok P, de Lange FP. 2014. Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex. *Curr. Biol.* 24(13):1531–35

Kok P, Jehee JFM, de Lange FP. 2012. Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron* 75(2):265–70

Kourtzi Z, Welchman AE. 2019. Learning predictive structure without a teacher: decision strategies and brain routes. *Curr. Opin. Neurobiol.* 58:130–34

Kriegeskorte N. 2015. Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annu. Rev. Vis. Sci.* 1:417–46

Kuai S-G, Zhang J-Y, Klein SA, Levi DM, Yu C. 2005. The essential role of stimulus temporal patterning in enabling perceptual learning. *Nat. Neurosci.* 8(11):1497–99

Kubilius J, Bracci S, Op de Beeck HP. 2016. Deep neural networks as a computational model for human shape sensitivity. *PLOS Comput. Biol.* 12(4):e1004896

Lake BM, Baroni M. 2018. Generalization without systematicity: on the compositional skills of sequence-to-sequence recurrent networks. In *Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden*, pp. 2873–82. N.p.: PMLR

Lake BM, Salakhutdinov R, Tenenbaum JB. 2015. Human-level concept learning through probabilistic program induction. *Science* 350(6266):1332–38

Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ. 2017. Building machines that learn and think like people. *Behav. Brain Sci.* 40:e253

Law C-T, Gold JI. 2008. Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nat. Neurosci.* 11:505–13

Lee ALF, Liu Z, Lu H. 2021. Parts beget parts: bootstrapping hierarchical object representations through visual statistical learning. *Cognition* 209:104515

Lee TS, Mumford D. 2003. Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A* 20(7):1434–48

LeMessurier AM, Feldman DE. 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. *J. Vis.* 19(4):28

Lengyel G, Nagy M, Fiser J. 2021. Statistically defined visual chunks engage object-based attention. *Nat. Commun.* 12:272

Lengyel G, Žalalytė G, Pantelides A, Ingram JN, Fiser J, et al. 2019. Unimodal statistical learning produces multimodal object-like representations. *eLife* 8:e43942

Li W. 2016. Perceptual learning: use-dependent cortical plasticity. *Annu. Rev. Vis. Sci.* 2:109–30

Luo Y, Zhao J. 2018. Statistical learning creates novel object associations via transitive relations. *Psychol. Sci.* 29(8):1207–20

Ma WJ, Beck JM, Latham PE, Pouget A. 2006. Bayesian inference with probabilistic population codes. *Nat. Neurosci.* 9(11):1432–38

MacKenzie K, Fiser J. 2010. Sensitivity of implicit visual rule-learning to the saliency of the stimuli. *J. Vis*. 8:474

Maniglia M, Seitz AR. 2018. Towards a whole brain model of perceptual learning. *Curr. Opin. Behav. Sci.* 20:47–55

Marcus GF, Johnson S, Fernandes K, Slemmer J. 2004. *Rules, statistics and domain-specificity:*

evidence from prelinguistic infants. Paper presented at the 29th Annual Meeting of the Boston University Conference on Language Development, Nov. 5–7. https://www.bu.edu/bucld/files/2011/06/handbook-292004.pdf

Marcus GF, Vijayan S, Bandi Rao S, Vishton PM. 1999. Rule learning by seven-month-old infants. *Science* 283(5398):77–80

Mareschal D, French RM. 2017. TRACX2: a connectionist autoencoder using graded chunks to model infant visual statistical learning. *Philos. Trans. R. Soc. Lond. B* 372(1711):20160057

Mark S, Moran R, Parr T, Kennerley SW, Behrens TEJ. 2020. Transferring structural knowledge across cognitive maps in humans and models. *Nat. Commun.* 11:4783

Marr D. 1982. *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. New York: Freeman

Maunsell JHR. 2015. Neuronal mechanisms of visual attention. *Annu. Rev. Vis. Sci*. 1:373–91

Minda JP, Smith JD. 2001. Prototypes in category learning: the effects of category size, category structure, and stimulus complexity. *J. Exp. Psychol. Learn. Mem. Cogn.* 27(3):775–99

Murphy RA, Mondragon E, Murphy VA. 2008. Rule learning by rats. *Science* 319(5871):1849–51

Murray RF. 2021. Lightness perception in complex scenes. *Annu. Rev. Vis. Sci.* 7:417–36

Murray SO, Kersten D, Olshausen BA, Schrater P, Woods DL. 2002. Shape perception reduces activity in human primary visual cortex. *PNAS* 99(23):15164–69

Musz E, Weber MJ, Thompson-Schill SL. 2015. Visual statistical learning is not reliably modulated by selective attention to isolated events. *Atten. Percept. Psychophys.* 77(1):78–96

Nadasdy Z, Nguyen TP, Török Á, Shen JY, Briggs DE, et al. 2017. Context-dependent spatially periodic activity in the human entorhinal cortex. *PNAS* 114(17): E3516–25

Nemeth D, Janacsek K, Londe Z, Ullman MT, Howard DV, Howard JH. 2010. Sleep has no critical role in implicit motor sequence learning in young and old adults. *Exp. Brain Res.* 201:351–58

Niv Y. 2019. Learning task-state representations. *Nat. Neurosci.* 22(10):1544–53

O'Keefe J, Dostrovsky J. 1971. The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* 34:171–75

O'Keefe J, Nadel L. 1978. *The Hippocampus as a Cognitive Map*. Oxford, UK: Oxford Univ. Press

Ongchoco J, Uddenberg S, Chun M. 2016. Statistical learning of movement. *J. Vis*. 16:1079

Orbán G, Berkes P, Fiser J, Lengyel M. 2016. Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron* 92(2):530–43

Orbán G, Fiser J, Aslin RN, Lengyel M. 2008. Bayesian learning of visual chunks by human observers. *PNAS* 105(7):2745–50

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* 147:113–26

Overlan MC, Jacobs RA, Piantadosi ST. 2017. Learning abstract visual concepts via probabilistic program induction in a language of thought. *Cognition* 168:320–34

Palmer S, Rock I. 1994. Rethinking perceptual organization: the role of uniform connectedness. *Psychon. Bull. Rev.* 1:29–55

Peña M, Bonatti LL, Nespor M, Mehler J. 2002. Signal-driven computations in speech processing. *Science* 298(5593):604–7

Perruchet P. 2019. What mechanisms underlie implicit statistical learning? Transitional probabilities versus chunks in language learning. *Topics Cogn. Sci.* 11(3):520–35

Perruchet P, Pacton S. 2006. Implicit learning and statistical learning: one phenomenon, two approaches. *Trends Cogn. Sci.* 10(5):233–38

Perruchet P, Vinter A. 1998. PARSER: a model for word segmentation. *J. Mem. Lang*. 39(2):246–63

Pinker S, Jackendoff R. 2005. The faculty of language: What's special about it? *Cognition*

95(2):201–36

Plaut DC, Vande Velde AK. 2017. Statistical learning of parts and wholes: a neural network approach. *J. Exp. Psychol. Gen.* 146(3):318–36

Pomerantz JR, Sager LC, Stoever RJ. 1977. Perception of wholes and of their component parts: some configural superiority effects. *J. Exp. Psychol. Hum. Percept. Perform.* 3(3):422–35

Pouget A, Beck JM, Ma WJ, Latham PE. 2013. Probabilistic brains: knowns and unknowns. *Nat. Neurosci.* 16(9):1170–78

Pouncy T, Tsividis P, Gershman SJ. 2021. What is the model in model-based planning? *Cogn. Sci.* 45:e12928

Rabagliati H, Ferguson B, Lew-Williams C. 2019. The profile of abstract rule learning in infancy: meta-analytic and experimental evidence. *Dev. Sci.* 22:e12704

Rabi R, Minda JP. 2014. Rule-based category learning in children: the role of age and executive functioning. *PLOS ONE* 9:e85316

Radulescu A, Shin YS, Niv Y. 2021. Human representation learning. *Annu. Rev. Neurosci.* 44:253–73

Retailleau A, Morris G. 2018. Spatial rule learning and corresponding CA1 place cell reorientation depend on local dopamine release. *Curr. Biol.* 28(6):836–46.e4

Riesenhuber M, Poggio T. 1999. Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2:1019–25

Roelfsema PR, de Lange FP. 2016. Early visual cortex as a multiscale cognitive blackboard. *Annu. Rev. Vis. Sci.* 2:131–51

Rosa-Salva O, Fiser J, Versace E, Dolci C, Chehaimi S, et al. 2018. Spontaneous learning of visual structures in domestic chicks. *Animals* 8(8):135

Rosch E. 1973. Natural categories. *Cogn. Psychol.* 4(3):328–50

Rosch E. 1975. Cognitive representations of semantic categories. *J. Exp. Psychol. Gen.* 104(3):192–233

Saffran JR, Aslin RN, Newport EL. 1996. Statistical learning by 8-month-old infants. *Science* 274(5294):1926–28

Saffran JR, Hauser M, Seibel R, Kapfhamer J, Tsao F, Cushman F. 2008. Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition* 107(2):479–500

Saffran JR, Kirkham NZ. 2018. Infant statistical learning. *Annu. Rev. Psychol.* 69:181–203

Saffran JR, Pollak SD, Seibel RL, Shkolnik A. 2007. Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition* 105(3):669–80

Sagi D. 2011. Perceptual learning in vision research. *Vis. Res.* 51(13):1552–66

Sagi D, Tanne D. 1994. Perceptual learning: learning to see. *Curr. Opin. Neurobiol.* 4(2):195–99

Santolin C, Rosa-Salva O, Vallortigara G, Regolin L. 2016. Unsupervised statistical learning in newly hatched chicks. *Curr. Biol.* 26(23):R1218–20

Schapiro AC, Turk-Browne NB, Botvinick MM, Norman KA. 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* 372(1711):20160049

Schonberg C, Marcus GF, Johnson SP. 2018. The roles of item repetition and position in infants' abstract rule learning. *Infant Behav. Dev.* 53:64–80

Schoups A, Vogels R, Orban GA. 1995. Human perceptual learning in identifying the oblique orientation: retinotopy, orientation specificity and monocularity. *J. Physiol.* 483(Pt. 3):797–810

Schoups A, Vogels R, Qian N, Orban G. 2001. Practising orientation identification improves orientation coding in V1 neurons. *Nature* 412(6846):549–53

Schulz E, Franklin NT, Gershman SJ. 2020. Finding structure in multi-armed bandits. *Cogn. Psychol.* 119:101261

Semedo JD, Zandvakili A, Machens CK, Yu BM, Kohn A. 2019. Cortical areas interact through a communication subspace. *Neuron* 102(1):249–59.e4

Sherman BE, Turk-Browne NB. 2020. Statistical prediction of the future impairs episodic encoding of the present. *PNAS* 117(37):22760–70

Siegelman N, Bogaerts L, Armstrong BC, Frost R. 2019. What exactly is learned in visual statistical learning? Insights from Bayesian modeling. *Cognition* 192:104002

Siegelman N, Bogaerts L, Elazar A, Arciuli J, Frost R. 2018. Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition* 177:198–213

Smith FW, Muckli L. 2010. Nonstimulated early visual areas carry information about surrounding context. *PNAS* 107(46):20099–103

Solway A, Diuk C, Córdova N, Yee D, Barto AG, et al. 2014. Optimal behavioral hierarchy. *PLoS Comput. Biol*. 10(8):e1003779

Sotiropoulos G, Seitz AR, Seriès P. 2011. Changing expectations about speed alters perceived motion direction. *Curr. Biol.* 21(21):R883–84

Srivastava S, Ben-Yosef G, Boix X. 2019. Minimal images in deep neural networks: fragile object recognition in natural images. arXiv:1902.03227 [cs.CV]

Tan Q, Wang Z, Sasaki Y, Watanabe T. 2019. Category-induced transfer of visual perceptual learning. *Curr. Biol.* 29(8):1374–78.e3

Tartaglia EM, Bamert L, Mast FW, Herzog MH. 2009. Human perceptual learning by mental imagery. *Curr. Biol.* 19(24):P2081–85

Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND. 2011. How to grow a mind: statistics, structure, and abstraction. *Science* 331(6022):1279–85

Tervo D, Gowanlock R, Tenenbaum JB, Gershman SJ. 2016. Toward the neural implementation of structure learning. *Curr. Opin. Neurobiol.* 37:99–105

Tolman EC. 1948. Cognitive maps in rats and men. *Psychol. Rev.* 55(4):189–208

Tomov MS, Schulz E, Gershman SJ. 2021. Multi-task reinforcement learning in humans. *Nat. Hum. Behav.* 5(6):764–73

Toro JM, Trobalón JB. 2005. Statistical computations over a speech stream in a rodent. *Percept.*

Psychophys. 67:867–75

Turk-Browne NB, Isola PJ, Scholl BJ, Treat TA. 2008. Multidimensional visual statistical learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 34(2):399–407

Turk-Browne NB, Jungé J, Scholl BJ. 2005. The automaticity of visual statistical learning. *J. Exp. Psychol. Gen.* 134(4):552–64

Ullman S, Assif L, Fetaya E, Harari D. 2016. Atoms of recognition in human and computer vision. *PNAS* 113(10):2744–49

van Bergen RS, Ma WJ, Pratte MS, Jehee JFM. 2015. Sensory uncertainty decoded from visual cortex predicts behavior. *Nat. Neurosci.* 18(12):1728–30

Vapnik VN. 1999. An overview of statistical learning theory. *IEEE Trans. Neural Netw*. 10(5):988–99

von der Heydt R, Peterhans E, Baumgartner G. 1984. Illusory contours and cortical neuron responses. *Science* 224(4654):1260–62

von Luxburg U, Schölkopf B. 2011. Statistical learning theory: models, concepts, and results. In *Handbook of the History of Logic*, ed. DM Gabbay, S Hartmann, J Woods, pp. 651–706. Amsterdam: Elsevier

Wagemans J, Elder JH, Kubovy M, Palmer SE, Peterson MA, et al. 2012. A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychol. Bull.* 138(6):1172–217

Wang JX. 2021. Meta-learning in natural and artificial intelligence. *Curr. Opin. Behav. Sci*. 38:90–95

Wang R, Wang J, Zhang J-Y, Xie X-Y, Yang Y-X, et al. 2016. Perceptual learning at a conceptual level. *J. Neurosci*. 36(7):2238–46

Wang R, Zhang J-Y, Klein SA, Levi DM, 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. 66:197–221

Wenliang LK, Seitz AR. 2018. Deep neural networks for modeling visual perceptual learning. *J. Neurosci.* 38(27):6028–44

Werchan DM, Amso D. 2020. Top-down knowledge rapidly acquired through abstract rule learning biases subsequent visual attention in 9-month-old infants. *Dev. Cogn. Neurosci.* 42:100761

Werchan DM, Collins AGE, Frank MJ, Amso D. 2015. 8-Month-old infants spontaneously learn and generalize hierarchical rules. *Psychol. Sci.* 26(6):805–15

Woods KJP, McDermott JH. 2018. Schema learning for the cocktail party problem. *PNAS* 115(14):E3313–22

Wu CM, Schulz E, Garvert MM, Meder B, Schuck NW. 2020. Correction: similarities and differences in spatial and non-spatial cognitive maps. *PLOS Comput. Biol.* 16(10):e1008384

Wu CM, Schulz E, Speekenbrink M, Nelson JD, Meder B. 2018. Generalization guides human exploration in vast decision spaces. *Nat. Hum. Behav.* 2(12):915–24

Xiao L-Q, Zhang J-Y, Wang R, Klein SA, Levi DM, Yu C. 2008. Complete transfer of perceptual learning across retinal locations enabled by double training. *Curr. Biol.* 18(24):1922–26

Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ. 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *PNAS* 111(23):8619–24

Yang S, Bill J, Drugowitsch J, Gershman SJ. 2021. Human visual motion perception shows hallmarks of Bayesian structural inference. *Sci. Rep.* 11:3714

Yildirim I, Jacobs RA. 2013. Transfer of object category knowledge across visual and haptic modalities: experimental and computational studies. *Cognition* 126(2):135–48

Yu C, Klein SA, Levi DM. 2004. Perceptual learning in contrast discrimination and the (minimal) role of context. *J. Vis.* 4(3):169–82

Yuille A, Kersten D. 2006. Vision as Bayesian inference: analysis by synthesis? *Trends Cogn.*

Sci. 10(7):301–8

Zhang J-Y, Kuai S-G, Xiao L-Q, Klein SA, Levi DM, Yu C. 2008. Stimulus coding rules for perceptual learning. *PLOS Biol.* 6(8):e197

Zhao J, Ngo N, McKendrick R, Turk-Browne NB. 2011. Mutual interference between statistical summary perception and statistical learning. *Psychol. Sci.* 22(9):1212–19

Zhou H, Friedman HS, von der Heydt R. 2000. Coding of border ownership in monkey visual cortex. *J. Neurosci.* 20(17):6594–611