Perceptual learning and representational learning in humans and animals

József Fiser
Brandeis University, Waltham, Massachusetts

Traditionally, perceptual learning in humans and classical conditioning in animals have been considered as two very different research areas, with separate problems, paradigms, and explanations. However, a number of themes common to these fields of research emerge when they are approached from the more general concept of representational learning. To demonstrate this, I present results of several learning experiments with human adults and infants, exploring how internal representations of complex unknown visual patterns might emerge in the brain. I provide evidence that this learning cannot be captured fully by any simple pairwise associative learning scheme, but rather by a probabilistic inference process called *Bayesian model averaging*, in which the brain is assumed to formulate the most likely chunking/grouping of its previous experience into independent representational units. Such a generative model attempts to represent the entire world of stimuli with optimal ability to generalize to likely scenes in the future. I review the evidence showing that a similar philosophy and generative scheme of representation has successfully described a wide range of experimental data in the domain of classical conditioning in animals. These convergent findings suggest that statistical theories of representational learning might help to link human perceptual learning and animal classical conditioning results into a coherent framework.

The phenomenon of learning is central to humans and animals, and as such it is the focus of an intense scientific investigation. However, due to the complex nature of learning, the areas of research on learning are multiple, with little crosslinking between these areas, even in the more restricted domain of behavioral and system neuroscience. The goal of the present review is to highlight some recent empirical and theoretical developments in both human and animal learning that suggest a possibly more general framework within which learning could be interpreted and investigated successfully. The more general framework is grounded in a statistical reformulation of different learning theories; more specifically, expressing learning as an implicit Bayesian inference carried out by humans and animals.

The structure of the article is as follows. First, I describe the classical definition of human perceptual learning and the way this classical paradigm could be modified a representational learning paradigm, to examine how humans acquire new complex representations of their environment. Next, I explain a number of results in human learning within this representational learning paradigm, and I show how these results are naturally explained within a Bayesian framework. In the last part of the article, I shift gears and argue that classical conditioning in animals is a sort of perceptual learning and can be reformulated in a representational learning paradigm, in much the same way as classical perceptual learning in humans can be. I present the results of studies that reformulated represen-

tational paradigm of classical conditioning in a Bayesian framework and thus afforded better insight into the underlying computation than did classical models. In the concluding part of the article, I highlight parallels between the two approaches and propose how representational learning might shed new light on the computational nature of a large class of learning phenomena.

CLASSICAL PERCEPTUAL LEARNING IN HUMANS

Perceptual learning has traditionally been defined as a practice-induced improvement in humans' ability to perform specific perceptual tasks (Fahle & Poggio, 2002). In perceptual learning paradigms, the subject is presented with a well-defined task explained verbally by the experimenter—orientation discrimination, for example (Fiorentini & Berardi, 1980; Furmanski & Engel, 2000; Petrov, Dosher, & Lu, 2006), texture discrimination (Ahissar & Hochstein, 1997; Karni & Sagi, 1991), motion direction discrimination (Matthews, Liu, Geesaman, & Qian, 1999), or a hyperacuity test (Poggio, Fahle, & Edelman, 1992). After repetitive training (typically including feedback), the subject's performance improves as quantified by threshold or reaction time measures. This experimental paradigm has been explored extensively in both the psychophysical (Dosher & Lu, 1998; Furmanski & Engel, 2000; Gold, Bennett, & Sekuler, 1999) and the neurophysiological domains (Gilbert, Sigman, & Crist,

J. Fiser, fiser@brandeis.edu

2001; Schoups, Vogels, Qian, & Orbán, 2001), and it is considered the dominant experimental approach to human sensory learning that does not strictly deal with abstract cognitive tasks such as concept learning (Fahle & Poggio, 2002; Fine & Jacobs, 2002).

Despite this widespread view, the classical paradigm captures only one type of change occurring in the brain due to controlled and limited sensory experience during fine discrimination tasks (Goldstone, 1998). However, learning is a much more complex phenomenon (Gallistel, 1990), and if the focus of the investigation is on understanding how humans and animals adaptively react to their complex surroundings in the most efficient way, a different aspect of human learning becomes important. In the visual modality, this concerns how infants learn their visual environment, the existence of objects and their interactions in scenes (Kellman & Arterberry, 1998), and the process by which adults acquire new visual knowledge of visual inputs never before seen (Kemp & Tenenbaum, 2008). Although this type of learning is based on perceived sensory input, and thus can be called *perceptual learning*, clearly it focuses on a different aspect of learning than traditional perceptual learning studies do. The present review will focus on such learning in the visual domain, but I suggest that the conclusions will generalize to such learning based on any modality.

In vision, the question of how to learn to understand our environment can be restated as how humans and animals extract novel visual descriptions, features, or chunks from their visual environment and represent them so that they can later use them for the purposes of effective visual recognition and recognition-based action. I suggest that there are a number of essential characteristics of this process that set it apart from the tasks typically used in traditional studies of perceptual learning. First, it is implicit or unsupervised, in that there is no explicit external direction as to what should or should not be retained from the visual input. Second, it works on an extremely complex, hierarchically structured, spatiotemporal input, where not only do many embedded substructures compose the scene, but many of them participate in multiple contexts at different times. This requires a very sophisticated representation, by contrast to the very simple stimuli used in traditional perceptual learning studies. Third, the interactions between objects in the outside world can alter the appearance of these structures substantially; for successful application, their description must, therefore, provide a large degree of generalization. Fourth, since a simple brute-force approach of learning all the possible features cannot cope with the complexity of real visual inputs due to the explosive combinatorics of the problem, the proposed learning mechanism must be computationally powerful enough to solve the task and meanwhile remain biologically plausible.

These features of the problem show that the classically defined paradigm of visual perceptual learning is not very appropriate, for two reasons, to investigating how humans develop complex visual representations in the natural environment. First, whereas the task in representational learning is to automatically extract new descriptions from a large set of potential descriptions, during classical perceptual learning subjects improve their threshold of discrimination between well-specified simple patterns, which can be done by increasing the sensitivity of existing detectors. Second, whereas representational learning is largely detached from conscious access, the attribute to be learned during classical perceptual learning is prespecified in the task description, and, therefore, the learning process is much more influenced by cognitively controlled processes (but see Watanabe, Náñez, & Sasaki, 2001, and in this issue).

There are three aspects of classical perceptual learning that need to be altered for a new framework, termed the *observational learning paradigm*, to make it suitable for investigating how humans acquire new internal representations. First, although the overall goal of observational learning is to encode useful aspects of the low-level sensory input for further processing, as in the case of classical perceptual learning, the fundamental *computational* task is not that Gaussian noise obscures some of the elements of the input and hence prevents the correct perception of the input that could lead to learning of some useful underlying associations; rather, sensory input there is fairly clear but ambiguous. It can support far too many possible combinations of elements that all could be potentially relevant higher order features. This is a combinatorial problem and not a signal-processing problem that lies at the heart of classic perceptual learning studies; the new paradigm should therefore be suitable for testing this combinatorial aspect of the learning problem in a controllable way.

Second, as was mentioned above, traditional perceptual learning tasks use explicit feedback (although it is not always related to the specific attribute learned; cf. Seitz & Watanabe, 2003), so they test how humans acquire new information in a supervised situation, whereas the principal challenge of developing new representations is to provide an unsupervised method of learning. Supervised learning is a much simpler task than unsupervised, in the sense that there is a single well-defined objective function and, in each trial, very detailed error information is provided to retune the system. In contrast, the only objective function in unsupervised learning is to capture the structure of the input. Even though it is clear that humans perform both unsupervised and supervised learning, and that there is a strong interaction between the two, the fundamental level of this process (translating light information into meaningful visual interpretations) is better characterized by unsupervised learning. Although many processes of human knowledge acquisition are goal directed and also rely on explicit external error measures, the first step toward these more cognitive types of memory formations is a dimension-reducing unsupervised learning process. This process, based on the external structure of the visual world, is intended to develop a representation that supports higher level learning; it proceeds without a bias imposed by an explicit task. To investigate this process, we need an experimental learning paradigm that is inherently unsupervised.

The third aspect is related to the stimuli used in perceptual learning tasks. Since these studies focus on the general mechanisms of visual learning, the stimuli need to be

designed so as not to contaminate learning through some uncontrolled specific features. Simple visual stimuli can form intermediate-level features by recourse to already existing lower level "grouping principles," and the internal representation on the basis of which learning occurs thus becomes inaccessible to the experimenter. For example, although classical perceptual learning has been studied for decades, it is still not known what exactly is learned when the subject can perform the orientation discrimination task better. Paradoxically, to have total control over all statistical features of the stimulus that the subject can utilize for learning is not always accomplished by using the simplest stimuli. Even though displays with a number of localized Gabor patches would be the type of fundamental stimuli traditionally used in the visual psychophysics of perceptual learning, they are very similar, and this causes a problem. When only a few Gabor patches are used in each display, the visual input is too simple to provide the necessary environment for studying statistical learning. Conversely, when more than a few Gabor patches are shown in the display, the visual system immediately invokes a number of previously developed midlevel representations based on configural or grouping mechanisms. For example, a cluster of semicontinuously oriented Gabor patches will "pop out" of a background array of randomly oriented Gabor patches (Kovacs, Kozma, Fehér, & Benedek, 1999). Thus, the statistical learning mechanism will work on these intermediate representations rather than on the structures controlled explicitly by the experimenter. The modified observational learning paradigm needs to avoid this pitfall and work with stimuli whose relevant statistical features are completely controlled.

FROM PERCEPTUAL LEARNING TO REPRESENTATIONAL LEARNING

What new learning paradigm could be developed to handle the challenges listed in the preceding section? The most fitting existing paradigm for observational learning is implicit learning—more specifically, statistical learning. Implicit learning is the ability to learn without awareness and has been explored mostly in language acquisition and skill learning research (Stadler & Frensch, 1997). Although implicit learning is thought to be incidental but robust, there is no agreement in the field regarding the extent to which it produces abstract complex knowledge (Cleeremans, Destrebecqz, & Boyer, 1998). Since early implicit learning studies were conducted in the domain of language acquisition, "abstract complex knowledge" has been defined initially as "knowing the grammar" that produces a set of observed strings (Reber, 1967). In implicit learning studies conducted in other domains, this concept of abstraction was generalized to mean "knowing the rules" that produced the observed inputs (Lewicki, Hill, & Bizot, 1988). However, recent studies have proposed that using a grammar or a rule efficiently might not even require explicit knowledge of the rules themselves (Perruchet & Pacton, 2006). It might be sufficient to use a simple associative learning mechanism that can trace distributional properties of the input (by

statistical learning) or a chunking mechanism that finds small typical fragments of the input (by chunk learning) to account for all the rule-learning results of implicit learning experiments.

Statistical learning is a special version of implicit learning that, rather than invoking symbolic rules, focuses on humans' ability to learn the simple statistical structures of the input. Statistical learning was first also introduced in the domain of language acquisition (Saffran, Aslin, & Newport, 1996; Saffran, Newport, Aslin, Tunick, & Baruecco, 1997). These studies showed that mere exposure to a continuous sequence of auditory syllables, visual shapes, or full visual scenes is sufficient for adults and infants to extract regularities from the input. Specifically, in the Saffran et al. studies, subjects incidentally learned the transitional probabilities between syllables presented in a continuous uninterrupted auditory stream generated by a certain rule; that is, they automatically learned which syllable was most likely to follow a given syllable. Statistical learning has been extended to the domain of touch (Conway & Christiansen, 2006; Hunt & Aslin, 2001) and even to other species (Hauser, Newport, & Aslin, 2001), suggesting that it is a very general domain-independent behavioral phenomenon. Since statistical learning is unsupervised and implicit, it is uniquely adequate for observational learning, in both adults and infants. Indeed, this is the paradigm that we used in the modality of vision to make it suitable for studying human representational learning in both adults and infants (Fiser & Aslin, 2001, 2002a, 2002b, 2005).

Briefly, we used arbitrary configurations of complex, highly discriminable novel shapes as the elements, and generated visual scenes by combining a subset of these shapes in each scene, according to some statistical rules. The general properties of statistical learning could be investigated just as well by using these complex shapes as by using Gabor patches or other simple stimuli, but without interference from perceptual mechanisms already in place. Moreover, by randomly assigning shapes across subjects, we could eliminate any specific effect of perceptual mechanisms already in place for the perception of the shape elements that would be based on peculiar low-level feature co-occurrences, and thus we could very precisely control the relevant statistics for learning. The paradigm is completely unsupervised and implicit, and it tests humans' ability to extract higher order statistical regularities from unknown inputs. Although these experiments used compositions of simple black shapes as stimuli, they clarified the computational basis of human feature learning, and thus opened the road for nonclassical perceptual learning experiments with more realistic stimuli that can directly target the issue of humans' internal visual representation of the external world and how it develops.

FEATURES OF VISUAL REPRESENTATIONAL LEARNING

The following results demonstrate not only the advantages of the proposed learning paradigm, but also how these investigations can be tied to computational models
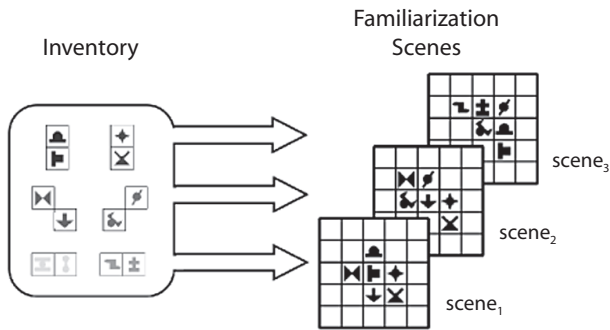
Figure 1. Generating familiarization scenes. Left: Twelve shapes organized into six pair combos with particular spatial arrangements constitute the inventory. Right: Sample scenes generated from the combos using three randomly selected combos at a time and arranging them randomly on a 5 × 5 grid.

to gain further insights into the nature of human visual representational learning.

Humans Are Automatically Sensitive to "Suspicious Coincidences" From Birth

All the subsequent human visual learning experiments followed the same basic design. Typically, we used 12 black shapes on a white background and created an inventory of structured building blocks called *combos*. A combo was a set of 2 or more shapes in a particular spatial (or temporal) configuration; any time one element of the combo appeared, all other elements of the combo also appeared in the predefined spatial (or temporal) configuration. The inventory of combos was then used to generate a large number of scenes by selecting a number of combos randomly for each scene and placing them on a grid in various adjacent configurations, so that none of the combos was completely separated from the other combos in the scene (Figure 1). The resulting scenes contained 6 or more shapes shown on a rectangular grid, so that there were no grouping cues for shape combos except the statistical cooccurrence of shapes across scenes.

An experiment was then conducted in two phases. In the first familiarization phase, subjects saw a long sequence of multishape scenes, each 2 sec long with a 1-sec pause between them. The subjects had no explicit task other than paying attention to the scenes, and they received no feedback of any kind. In the second testing phase, a set of two-alternative forced choice (2AFC) trials were given to the subject. In each trial, two displays were presented for 1 sec, of which one showed a single combo or an embedded part of a combo and the other showed randomly selected individual shapes arranged in combo format (Figure 2). An embedded part of a combo having three or more elements was a subset of those elements, appearing in the same arrangement as in the combo (see Figure 2, bottom panel). The random combination of elements showed shapes in an arrangement that they had never appeared in during familiarization. The subject had to choose the arrangement that looked more familiar, on the basis of the prior familiarization scenes. Each subject's preference for the combo, or combo part, over the random

shape combination was taken as evidence that the subject had extracted (unconsciously and automatically) the underlying structure of the scenes.

Using this experimental design, we ran a large number of visual statistical learning experiments in the last couple of years. The goal of the first set of experiments was to establish whether humans are sensitive to second-order joint and conditional visual statistics in the spatial and temporal domains, and, if so, how universal this sensitivity is across adults and infants. This question is significant because it has long been established that efficient learning of internal representations of highly complex input is possible only if humans are capable of extracting such statistics (Barlow, 1989, 1990); but whether humans are truly capable of doing this has never been tested empirically.

First, we ran the simplest spatial version of the experiment with adults, using the inventory of six combo pairs (144 different six-shape scenes) and testing for pairs of shapes (Fiser & Aslin, 2001). We found that after 6–8 min of continuous exposure to the familiarization scenes, adults chose the base combos over the random pairing of shapes significantly more often. This was true whether or not the random pairs were composed so that the individual components appeared in the absolute positions they could appear in during familiarization. Specifically, in Experiment 1, subjects could rely on both shape co-occurrences and shape–position associations, since the elements of the random pairs never appeared together or in the given position
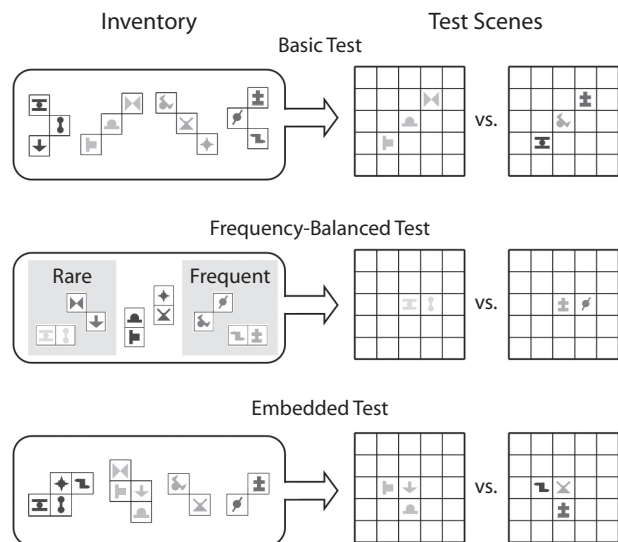

Figure 2. Three types of test used in the visual learning experiments. Left: A typical inventory of the given test type. Right: A typical example of a test trial. Top: A true triplet combo compared with a random triplet. All individual shapes appeared an equal number of times during practice. Middle: A true pair combo compared with a pair of elements selected from frequent combos. The elements of the two pairs co-occurred an equal number of times, but elements of the true combo always appeared together, whereas elements of the pair in the right could equally often appear without each other. Bottom: A part of a true quad combo compared with a random triplet. All individual shapes appeared an equal number of times during practice.
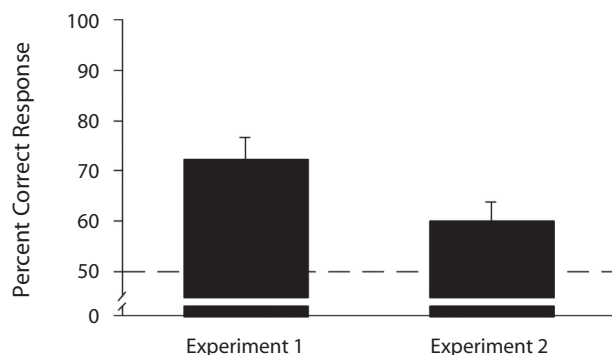
Figure 3. The results of the first two visual statistical learning experiments. Bars show percentages of responses in which the true pair combos were selected. Error bars show *SEM*s; 50% is chance.

during familiarization. In Experiment 2, only the number of shape cooccurrences differed between the true combo and the random pair. Both tests showed a significant preference for true combos, and there was a significant difference between the results of the two experiments. As assessed by a postexperimental interview, subjects were unaware of their performance or even of the underlying structure of the scenes they saw. This means that the subjects became sensitive to both the cooccurrence of shapes within the combos and the correlation between shapes and their position on the grid (Figure 3). These results confirmed that humans automatically and implicitly extract statistical structures of their visual experience, and that this representation relies on all aspects of the stimuli, including cooccurrences of and spatial information for individual elements.

In a third experiment, we used different appearance frequencies for different combos during training in order to create a situation where the cooccurrence of shapes within some combos was equated with the accidental cooccurrence of two shapes that did not belong to the same combo. With this *frequency-balanced* test, we could show that humans preferentially encode combos with two elements perfectly predicting each others' appearance, as compared with a pair of shapes that appeared together the same number of times as the true combo but that could also appear without each other (Figure 4). In other words, humans were sensitive to the conditional probability statistic of element pairs, even when the joint probabilities were equated. This provided an empirical support to Barlow's theory that human learning automatically detects "suspicious coincidences" in the environment and represents those as significant features (Barlow, 1989, 1990). We also found that, apart from the conditional probabilities between shapes, humans maintain sensitivity to the frequency of the individual shapes, since they could readily select the shapes that appeared more often during the practice session (Figure 4). Sensitivity to conditional probabilities in the spatial modality amounts to sensitivity to transitional probabilities in the temporal modality; this has been demonstrated in the auditory domain before (Aslin, Saffran, & Newport, 1998). In a new set of experiments, we demonstrated the same ability to learn such

temporal statistical regularities in the visual domain (Fiser & Aslin, 2002a).

To determine whether this ability to extract statistics incidentally from visual scenes is a fundamental learning tool for humans at all ages, we tested whether these sensitivities were present in infants as well. Simplified versions of the experiments above were run on 8-month-old infants in a preferential looking paradigm (Fiser & Aslin, 2002b). Here, the combos were pairs of elements, but the scenes consisted of only three elements—a combo and a corresponding noise element that could appear in an arbitrary position around the combo. We found that after about 2 min of familiarization with the scenes, infants looked at the picture of a base pair (a combo of the inventory) significantly longer than they did at a randomly combined pair of shapes from two combos (Figure 5). This replicated the results of the first two adult experiments.

We then asked whether infants would also perform well (as adults did) with frequency-balanced pairs. Infants looked significantly longer at the original combos than at the frequency-balanced random shape pairs (Figure 6). Interestingly, we found that infants did not replicate the adult results from the single shape test; they were unable to distinguish between low-frequency and high-frequency shapes. Together, these results confirmed that human adults and infants have access to information about conditional probability relations in previously unknown visual scenes and thus, in principle, it is possible for them to extract higher order visual structures by statistical learning mechanisms.

Humans Encode Visual Information in a Minimally Sufficient Manner

In the next set of experiments, we turned to the well-known problem of the "curse of dimensionality" or "combinatorial explosion" of statistical learning (Bellman, 1961; von der Malsburg, 1995). Briefly, this problem arises from the fact that statistical learning of higher order structures in a complex environment is impossible, even theoretically, because the number of training examples needed for learning the right structures is prohibitively large. This hard limit on brute force learning is in direct contradiction to the claim that humans learn their internal visual representations by statistical learning. In order to
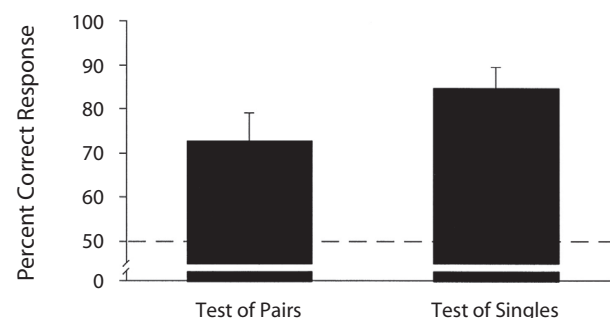

Figure 4. The results of the frequency-balanced experiment. Bars show percentages of responses in which the true pair combos were selected. Error bars show *SEM*s; 50% is chance.
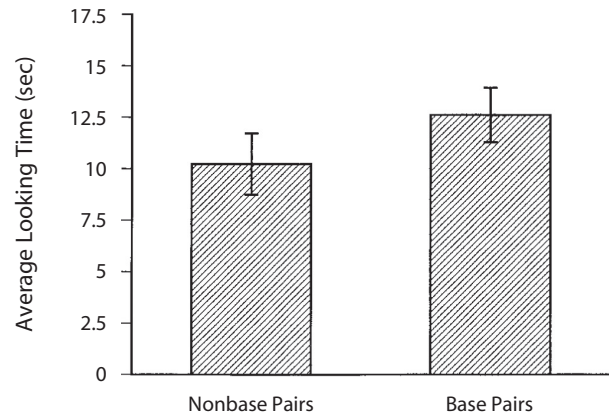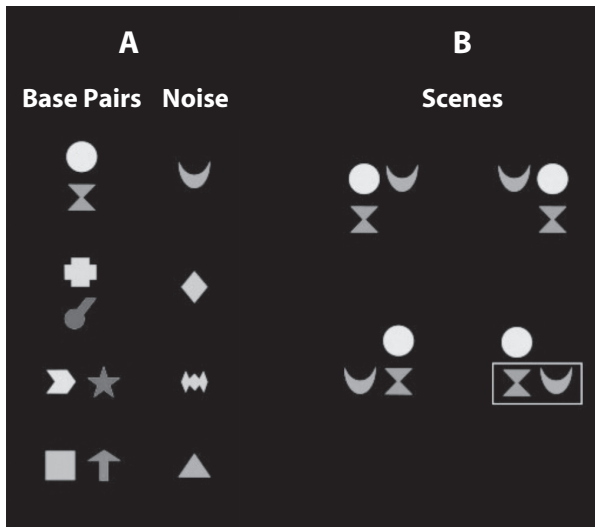
Figure 5. The infant visual statistical learning study. Left: (A) Inventory of the experiments with four combos (base pairs) and four noise elements. (B) Four possible scenes composed from one combo and its noise element. The white rectangle indicates the frequency-balanced pair used in the second pair test. Right: Results of the first pair test. Error bars show *SEM*s. Babies looked significantly longer at true combos.

investigate this issue, we used visual stimuli with hierarchical internal structure, where the number of available statistics for learning the underlying structure of the scene grew exponentially with the number of shapes involved in defining the underlying combos of the inventory (Fiser & Aslin, 2005). In the first experiment, we used the inventory of four triplet combos and tested whether humans would become sensitive to both the triplet structure and the structure of the pairs embedded in the triplet. We found that adults reliably selected the combo triplets over random shape triplets but that there was no such preference for the embedded pairs over random pairs (Figure 7, left).

In the next experiment, we increased the size of the combos and used the inventory of two quadruples and two pairs to construct the scenes. During the test, subjects had to choose between a quad combo and a random quad, a pair combo and random pairs, and an embedded pair and a random pair. Subjects preferred the quad and pair combos to the random structures, but they were at chance performance when they chose between a shape pair embedded in a quad and a random pair (Figure 7, right). This pattern did not change when we doubled the familiarization time. Interestingly, in a subsequent study, we found that embedded triplets of the same inventory were preferred over random triplets. In a series of control experiments, we ruled out a number of alternative explanations and concluded that this pattern of results suggests that subjects generate a minimally sufficient representation of combos instead of encoding the full structure of the underlying scenes (Fiser & Aslin, 2005). We also pointed out that that such a strategy could avoid the curse of dimensionality and fits naturally into a Bayesian framework.
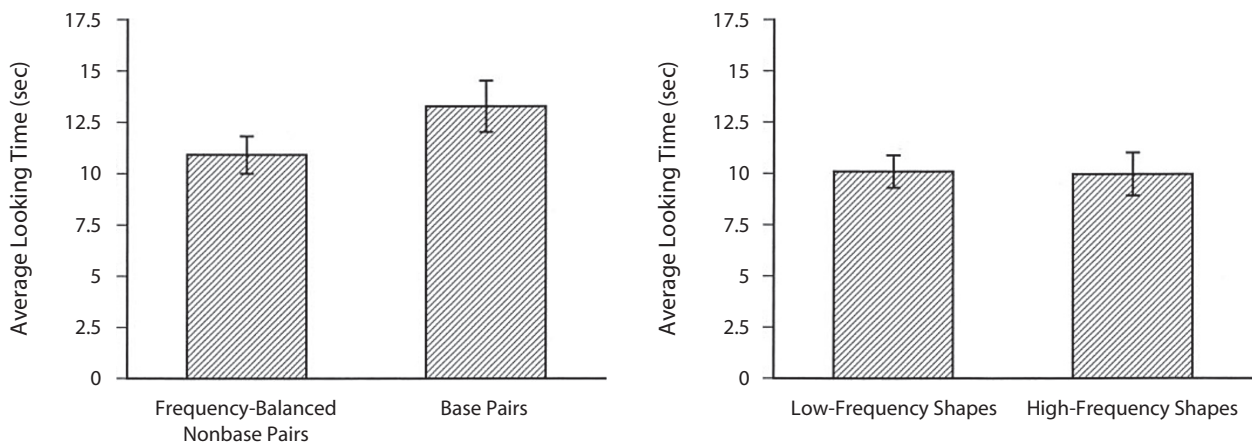

Figure 6. Results of the frequency-balanced infant learning experiment. Left: Results of the pair test. Babies looked longer at rare true combos. Right: Results of the single shape frequency test. There was no reliable difference in looking times between frequent and rare elements. Error bars show *SEM*s.
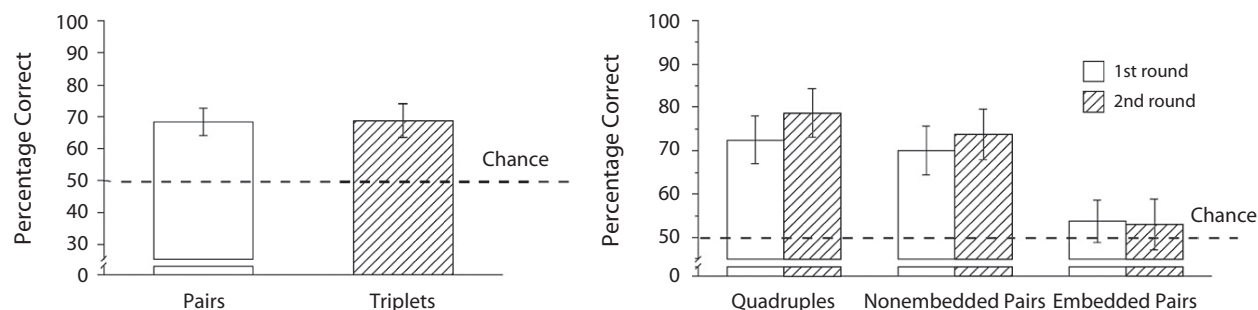
Figure 7. Results of the two embedded learning experiments. Left: Results of the embedded experiment with the inventory of triplet combos. Right: Results of the embedded experiment with the inventory of quadruple and pair combos. In both experiments, subjects reliably learned the true combos.

Humans Learn the Structure of Their Visual Environment in a Statistically Optimal Way

Using a Bayesian framework, we developed a computational model to explain all of our learning results, and we compared this model to other existing models of human unsupervised learning (Orbán, Fiser, Aslin, & Lengyel, 2008). These previous models can be grouped into roughly four types. Type 1 models are based on frequency counting, where the learning algorithm keeps track of the occurrence of individual patterns or events and patterns, with the highest frequencies representing the significant memory traces. This is the simplest type of learning model, and counting episodic memories belongs to this class. Type 2 models keep counts of pairwise cooccurrences of elements in the scenes. Apart from the important issue of how to define an element, this learning model is superior to Type 1 models because it operates on a vocabulary-based combinatorial description, not a single holistic representation. Type 3 models compute conditional or transitional probabilities between two elements of the scene rather than just their co-occurrence frequencies, and store combinations with high conditional probability as useful memory traces. Features obtained by this kind of learning are the "spurious coincidences" in the input stimulus explained by Barlow (1989, 1990, 2001). Type 3 models of learning are also the most popular among empirical researchers of statistical learning (Kuhl, 2004). In contrast to the first three models, Type 4 models are not simple event counters but rather full probabilistic models that learn all pairwise correlations between elements across scenes. The crucial difference between Type 3 and Type 4 models is that the latter assess the familiarity or likelihood of a new input on the basis of not only the elements that are present in the input, but also the elements absent from it. A Type 4 model is a statistically optimal implementation of the most extensively studied pairwise associative feature learning mechanisms in the literature that are typically modeled in unsupervised neural network architectures (Dayan & Abbott, 2001).

The model that we developed for testing against the previous models, which we refer to as *ideal learner*, is an explicit chunking model. Computationally, it is based on Bayesian model averaging, a statistically principled optimal way of solving the following problem: Given a set of atomic elements, how does one select an inventory of chunks, based on those elements, that makes it possible to capture the input in a minimally sufficient way and also to have the maximum ability to generalize to inputs never experienced before? In our model, each selected inventory (each choice of what combos or chunks will be the building blocks of scenes) defines a probability distribution over all possible scenes: $P(\text{scene}_1, \text{scene}_2, \ldots \text{scene}_n | \text{Inventory})$. Chunks are formalized as latent variables that describe the identity and relative positions of shapes making up the chunk. If the chunk is present in a scene, the probability of these shapes in the given configuration increases; if the chunk is absent, each element of the chunk can still appear with a "spontaneous" probability, regardless of what the other shapes do. Thus, when scenes are observed, chunks can be inferred as spurious coincidences of shapes, and for any set of chunks defining an inventory, the probability of familiarization scenes can be computed. On the basis of these probabilities, and assuming that chunks appear independently of one another, the best inventory can be selected by Bayesian model averaging by ranking the inventories according to the summed probability assigned by the inventory to the set of practice trials. Due to the "automatic Occam's razor" effect of Bayesian model averaging (MacKay, 2003), this method will select the optimal inventory that describes the previous practice scenes sufficiently well but does not prevent generalization to novel scenes.

We tested which of these five models best captures human performance in the learning experiments described above (Orbán et al., 2008). For this, we imported the training and test stimuli of each experiment, and trained and tested each model in exactly the same way as we did the humans. Figure 8 summarizes our findings. In the simplest test (panel A), the inventory contained six equal-frequency pairs. Only Type 1 failed to predict above-chance human performance on the basic test of true pairs versus mixture pairs. In the frequency balanced test (panel B), the inventory contained six pairs of varying frequency, and both Type 1 and Type 2 failed to predict above-chance human performance on the test of true rare pairs versus frequency-balanced mixture pairs. In the simple embedded experiment (panel C), the inventory contained four equal-frequency triplets, and human performance was above
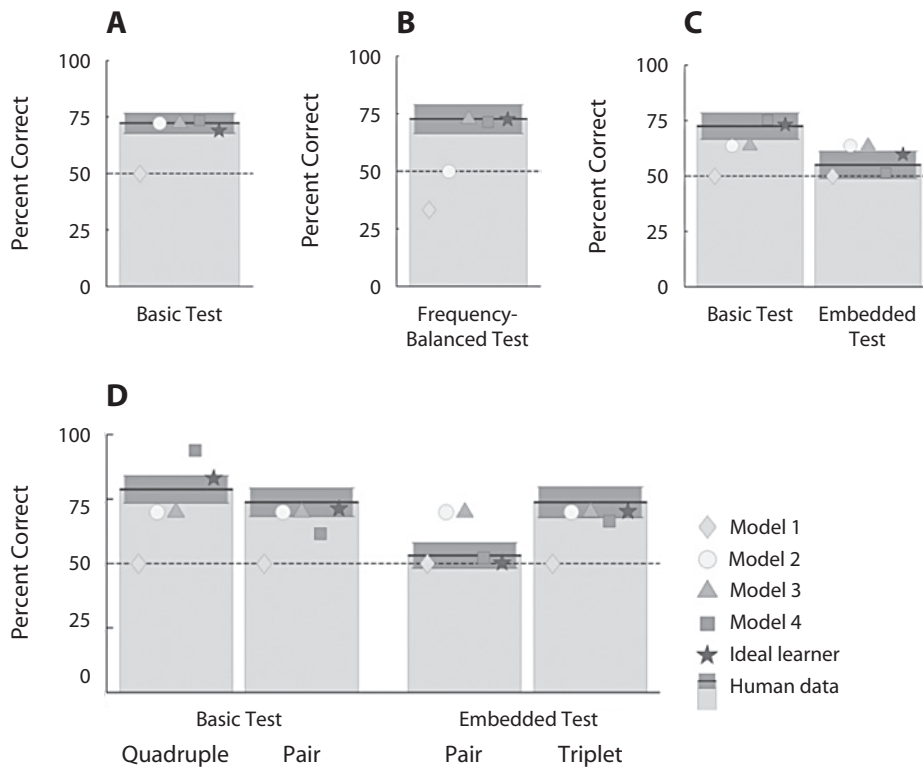
Figure 8. Summary of human performance (gray bars) and model predictions (colored symbols) for a series of experiments from Fiser and Aslin (2001, 2005a), using increasingly complex inventories. The black line and dark gray region at the top of each bar indicate mean and *SEM*, respectively.

chance on the basic test of true triplets versus mixture triplets, and at chance on the test of embedded pairs versus mixture pairs. Types 1, 2, and 3 incorrectly predicted the same performance on the basic and embedded tests. Finally, in the second embedded experiment (panel D), the inventory contained two quadruples and two pairs, all with equal frequency. Human performance was above chance on the basic tests of true quadruples or pairs versus mixture quadruples or pairs, and on the test of embedded triplets versus mixture triplets, but it was at chance on the test of embedded pairs versus mixture pairs. Types 1, 2, and 3 incorrectly predicted the same performance on all tests; only Type 4 and the ideal Bayesian learner captured the overall pattern of human performance in all these experiments. In summary, we found that only the ideal learner could qualitatively and quantitatively replicate all the patterns of human performance in all experiments.

However, the Type 4 model, most widely considered the appropriate model of human learning, performed relatively well; on the basis of these results, it cannot be discarded as a good model of human learning. Since Type 4 models and the ideal learner follow very different computational philosophies, however, we were able to design an experiment in which the predictions of the two models were contradictory and compare their results with human performance. Specifically, the inventory of the experiment had four so-called "circular triplets," four single elements that also sometimes formed a quadruple and two pairs (Fig-

ure 9). The key element of the design was that the four triplets shared shapes and the construction of the scenes was such that assessing only pairwise correlations was of no help in establishing the identity of the four circular triplets. Similarly, the four singletons appeared many times alone but also together as a quadruple, so that their pairwise correlations were also balanced. Thus, from the standpoint of first- and second-order correlations, there was no difference between the circular triplets and triplets composed from the singletons of the inventory. However, when higher order correlations are considered, the circular triplets were true building blocks of the inventory, whereas the singleton-based triplets were just faulty associations that did not capture the significant chunks in the scenes precisely enough.

We compared the prediction of the models with human performance in three tests: circular versus random triplets, singleton-based versus random triplets, and circular versus singleton-based triplets (Figure 10). Humans showed a significant preference for circular triplets in the first and third tests, and a chance performance in the second test. The pattern of predictions showed a clear difference between the Type 4 model and the ideal learner. We used the models, with their parameters tuned to account for the results in the previous experiments, to predict the model performances in the present experiment. The ideal learner followed human performance in all three experiments. In the first one, it correctly showed that circular triplets were
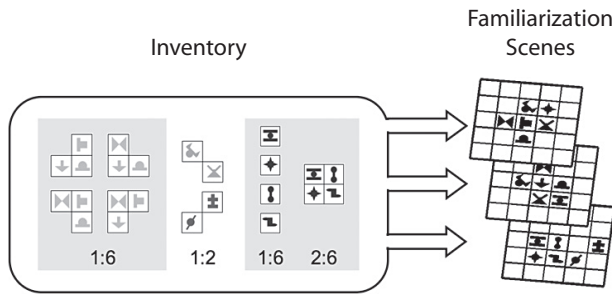
Figure 9. The inventory of the experiment testing the difference between Type 4 models and the ideal learner. Combos in the shaded areas represent the circular triplets and the singleton combos, respectively. Ratios at the bottom show the relative frequency of the combos across the practice scenes. All element appearance frequencies and pair co-occurrence frequencies were equated.

the true building blocks of the visual scenes. In the second test, it correctly signaled that a singleton-based triplet was not a very significant chunk of the visual scenes, despite having appeared across the set of scenes twice as often as a circular triplet. In the third, direct comparison, it significantly favored the circular triplets over singleton-based triplets. In stark contrast, the predictions of model directly contradicted human performance by completely equating the significance of the two types of test triplets. As a result, even though the model correctly favored the circular triplets over random ones in the first test, it favored the singleton-based triplets equally in the second test, and could not tell the difference between the two in the third direct test. These results suggest that, rather than learning all pairwise correlations of an input, humans follow optimal model averaging and achieve a minimally sufficient chunk-based representation of the input.

BAYESIAN APPROACHES TO CLASSICAL CONDITIONING IN ANIMALS

The preceding sections highlighted the distinct advantages of investigating human visual learning within a modified perceptual learning paradigm, which views the learning process as a complex representational problem

and approaches this problem from a statistical standpoint. In the following sections, I examine whether a similar approach can be applied to a seemingly very different learning domain: classical conditioning in animals.

Classical conditioning studies investigate how animals learn to predict a certain event (the unconditioned stimulus, or US), as measured by their conditioned response (CR), on the basis of reliably correlating available cues (the conditioned stimulus, or CS; see Pearce & Bouton, 2001). It is easy to see how this problem can be reformulated as a representational learning problem; the goal of learning is to develop a representation of the outside world including the relations between the CS and the US, and, given this acquired world model, to successfully infer the likelihood of the US under various conditions. These approaches use two main types of world models, discriminative and generative (Courville, Daw, & Touretzky, 2006). The discriminative world model focuses only on the USs, given the CSs: $P[US(t)\,|\,CS(t), D(t-1, \ldots 1)]$, where D represents all previous experiences. In contrast, the generative world model assumes that the animal develops an internal representation of the whole pattern of events, in the case of conditioning including both the CS and US, as $P[US(t), CS(t)\,|\,D(t-1, \ldots 1)]$, and uses this world model for predicting the likelihood of the US. It is clear that the generative approach (also called the *latent variable* approach) is more similar to the ideal Bayesian learner used for modeling human visual statistical learning in the preceding section.

Like human perceptual learning, the majority of models interpreting classical conditioning follow the associative approach (Rescorla & Wagner, 1972). These models correctly captured the earliest and simplest experimental results, but, since behavior under more elaborated experimental conditions has been identified in the large number of classical conditioning studies, the models capturing these new results also had to become more complex and varying (Mackintosh, 1975; Pearce & Hall, 1980). Bayesian models represent an alternative approach to explaining these results. On the basis of Courville et al. (2006), I describe four cases where an appropriately designed Bayesian model could provide useful insights into the nature of learning.
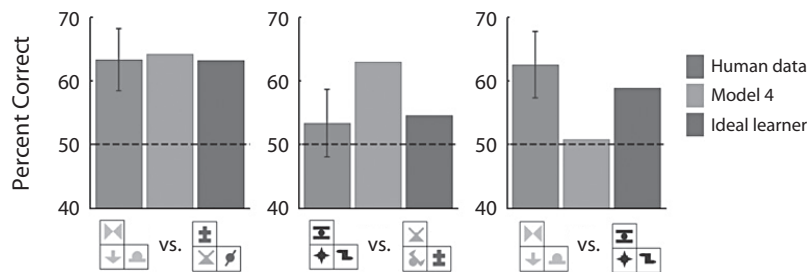

Figure 10. The results of the experiment testing the difference between Type 4 models and the ideal learner. The left panel shows the comparison between circular triplets and mixed triplets; the middle panel shows singleton-based triplets versus mixed triplets; and the right panel shows the direct comparison of circular triplets with singleton-based triplets. Only the ideal learner followed the pattern of human subjects' test performance.

Handling Similarity and Discrimination in Classical Conditioning

As in human perceptual learning, handling similarity between input patterns in order to generalize to new inputs is a cardinal issue in classical conditioning. Both the "elemental" associative model of Rescorla and Wagner (1972) and "configural" models with "configural units" (Pearce, 1994) can handle this issue, using different approaches (for a review, see Wagner, 2003). Courville, Daw, and Touretzky (2004) reformulated the model of classical conditioning in terms of Bayesian model averaging, a standard Bayesian method of accounting for the uncertainty that any particular model is the one that generated the data (Hoeting, Madigan, Raftery, & Volinsky, 1999). The advantage of this formalism is that it gives a principled way to determine the number and identity of the configural units, and to decide whether the Rescorla–Wagner or Pearce type of model captures the pattern of the data. Courville and his colleagues tested the performance of their model on three basic patterns of results of classical conditioning: overshadowing, summation, and overlap. They have demonstrated that Bayesian model averaging can reproduce these results as well as simple and configural methods can. However, a distinct advantage of the Bayesian approach is that it requires a complete model definition, and, once the model is defined, there is no further uncertainty about the selection of the configural units. In addition, there are cases when the Bayesian approach can explain results that the other two types of models cannot, as we will see below.

Learning the Temporal Structure of the Stimuli

R. R. Miller and colleagues conducted a series of experiments in the mid-1990s to investigate the extent to which animals learn the temporal structure of the stimuli (Barnet, Arnold, & Miller, 1991; Cole, Barnet, & Miller, 1995; Cole & Miller, 1999). The common theme of all these experiments is that they suggest a more complex representation of the events occurring during the training phase that involves encoding second-order conditioning and relative timing among multiple participating stimuli. Thus, the classical Rescorla–Wagner theory cannot predict these experimental outcomes, and neither can its more contemporary extensions (Sutton & Barto, 1990). What is needed, as suggested by Miller and colleagues' temporal coding hypothesis, is a compound representation of all the events, which representation lends itself to inferences about any relation between elements of this representation (Matzel, Held, & Miller, 1988). Courville and Touretzky (2002) noticed that this requirement is naturally fulfilled by Bayesian learning with a generative world model and developed such a computational model on the basis of a hidden Markov model that could successfully reproduce the results of all the experiments in the empirical studies mentioned above.

Handling Uncertainty of the Learned World Model

Even the classical Rescorla–Wagner model can be reformulated as a generative Bayesian model, but a very limited one, where only certain parameters of the model, such as the rate of the CS, can be learned, but the full model is fixed. The flexibility of a general Bayesian model allows expression of uncertainty about the contingencies by learning the right model as well as the right parameters of the model. Courville, Daw, Gordon, and Touretzky (2003) developed such a model and showed how it can be used to explain two peculiar results of classical conditioning, second-order conditioning and conditioned inhibition, that were not captured explicitly by earlier models of classical conditioning. Both in second-order conditioning and in conditioned inhibition, there are two types of trials: first, one in which a CS, A, appears with the US; and second, one in which A appears together with a different stimulus, C, only this time without any reinforcement. In second-order conditioning, then, C becomes an effective stimulus, since presenting it in the test trial elicits a CR. In contrast, in conditioned inhibition, C becomes an inhibitor during the test. More interestingly, the cause of this shift from exciting to inhibiting is due exclusively to an increase in the number of A–C trials during practice (Yin, Barnet, & Miller, 1994).

Bayesian model averaging can naturally accommodate these results on the basis of the same Occam's razor principle that was applied in the ideal Bayesian learner model on page 147. Specifically, having only a few A–C trials, there is not enough evidence available to promote a complex model to explain the data. The algorithm will therefore prefer a model with a single underlying cause and attribute both practice types to this cause. Thus, when C is tested alone, it is considered a "good enough" signal to assume the appearance of the US, leading to a CR. However, as evidence accumulates with more trials, a more complex model gains support, whereby A–US and A–C will be explained by two different causes; so, when C appears alone, it is evidence that the US will not appear, inhibiting the CR. Thus, including and handling uncertainty as to which model correctly captures the data leads to deeper understanding of the conditioning process. Courville et al. (2003) developed a Bayesian model selection scheme based on the ideas that matched the parameters and setup of the Yin et al. study and demonstrated how similarly the model and the animals behaved.

Handling Uncertainty Due to the Changing Nature of the World

A final example of the benefits of reformulating classical conditioning as a representational learning problem by using a Bayesian approach comes from investigating the effect of the changing environment on modeling animal behavior. Courville et al. (2006), used the Bayesian formalism to offer a conceptual explanation of the effect of change. The basic idea is to incorporate into the generative world model the possibility that the contingencies of the environment can change by letting the parameters of the model evolve with time. If the model assumes a changing environment, it will represent the speed of change, and other contingencies will depend on this information. Having such a representation in the model implies the notion that, in a changing world, animals will (possibly sub-

consciously) learn and represent the speed at which given contingencies change.

Once this modification to the world model is incorporated, the link between Bayesian models and classical associative models is straightforward. In Bayesian models, uncertainty about the inferred parameters is represented explicitly (Griffiths & Yuille, 2006), and this uncertainty is directly related to the associability defined in the associative models (Dayan & Long, 1998). Just as more uncertainty about a model parameter leads to more influence of newly obtained information (i.e., more change in the parameter) in the Bayesian model, larger associability leads to more learning in the associative model. The relation above also implies that uncertainty is directly linked to change. However, change is also directly related to surprise: When change in parameters occur, the new parameters will produce new events outside the scope of the old parameters—in other words, surprising events. Thus, if the world model incorporates representation of the speed of parameter changes, a surprising event will signal a rapid change in the parameter, which will in turn increase uncertainty in the parameter, giving rise to more rapid learning. This Bayesian account provides a clear formalism that links the different notions under a coherent view, all based on one empirically testable assumption: that animals develop internal representations of the speed of change in their environment.

Courville et al. (2006) implemented such a Bayesian model and tested it in a number of classical conditioning experiments. These included such experiments as latent inhibition (Lubow, 1973), reinforced preexposures (Hall & Pearce, 1979), and unblocking by reinforcer switch (Blaisdell, Denniston, & Miller, 1997), where surprising reinforcers speeded up learning, which was correctly captured by both the Pearce–Hall associative model (Pearce & Hall, 1980) and the Bayesian model. They also tested such experiments as latent inhibition and overshadowing (Blaisdell, Bristol, Gunther, & Miller, 1998), where a surprising neutral stimulus caused increased learning. These results could not be captured either by the Pearce–Hall associative model or by a Bayesian model that did not develop a representation of the world by a generative world model. Finally, Courville et al. (2006) used the partial reinforcement extinction effect (Gibbon, Farrell, Locurto, Duncan, & Terrace, 1980; Haselgrove, Aydin, & Pearce, 2004; Rescorla, 1999) to show that fixed-rate stochastic reinforcement might not induce faster learning. This is an interesting result, because the Pearce–Hall and the Bayesian models provide opposite predictions, and animal behavior follows the prediction of the Bayesian model.

DISCUSSION

In this article, I have reviewed two seemingly distinct areas of the research on learning, human perceptual learning and animal classical conditioning, that traditionally are not considered as dealing with similar issues. I argued that, despite obvious differences, the core question in these domains is essentially the same: How to combine the knowledge about observed perceptual events in order to react most effectively to future events? I proposed that this core question represents an extension of the classical definition of both domains that, nevertheless, is on the critical path toward a deeper understanding of the problem of learning and the computations in the brain that underlie solving this problem. I also suggested that the recently emerging probabilistic Bayesian framework, with a generative world model that is used to develop a full internal representation of the environment, is a natural candidate for investigating this question. On the basis of our research and that of others, I presented a number of examples to show that, in both domains, the Bayesian models can provide many times an equal or better capture of the empirical data, as well as new insights into why the system performs the computation it does. This is a further demonstration that placing the problem in a clear statistical framework helps to clarify the true nature of the computations performed by the brain.

A final interesting point emerging from these studies in the two domains is how Bayesian inference outperforms associative learning in given tasks. Associative learning has always been the leading candidate for the basic learning method by which the brain acquires new representations (Hebb, 1949). Clearly, detecting correlations among different aspects of the input signal must be the core of any learning. However, this by itself does not specify what representations will be developed on the basis of the correlations. Since the various options for solving the simple problem normalizing the Hebbian learning rule lead to very different final representations (Miller & MacKay, 1994), considering various aspects of the correlational structure in the input can lead to very different representations when those structures are prominent. In overly simple tasks, the simplest associative learning can perform very similarly to the more complex Bayesian inference based on internal representation. However, as shown on pages 148 and 151, when the complexity of the possible learning pattern increases, human and animal performance sharply diverge from what a simple associative scheme would predict. This suggests that the dynamic learning method implemented in the cortical circuitry of the brain performs a more sophisticated learning than simple associative learning does. The promise of the present article, and of a number of other recent studies investigating learning, is that a sufficiently sophisticated probabilistic learning model will be able to explain, under a unified computational principle, not only simple learning results such as classical perceptual learning and classical conditioning, but also more sophisticated types such as rule learning, reasoning, motor planning, and intelligent navigation.

REFERENCES

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

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, **9**, 321-324.

Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*, **1**, 295-311.

Barlow, H. B. (1990). Condition for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vision Research*, **30**, 1561-1571.

Barlow, H. B. (2001). Redundancy reduction revisited. *Network: Computation in Neural Systems*, **12**, 241-253.

Barnet, R. C., Arnold, H. M., & Miller, R. R. (1991). Simultaneous conditioning demonstrated in second-order conditioning: Evidence for similar associative structure in forward and simultaneous conditioning. *Learning & Motivation*, **22**, 253-268.

Bellman, R. (1961). *Adaptive control processes: A guided tour*. Princeton, NJ: Princeton University Press.

Blaisdell, A. P., Bristol, A. S., Gunther, L. M., & Miller, R. R. (1998). Overshadowing and latent inhibition counteract each other: Support for the comparator hypothesis. *Journal of Experimental Psychology: Animal Behavior Processes*, **24**, 335-351.

Blaisdell, A. P., Denniston, J. C., & Miller, R. R. (1997). Unblocking with qualitative change of unconditioned stimulus. *Learning & Motivation*, **28**, 268-279.

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

Cole, R. P., Barnet, R. C., & Miller, R. R. (1995). Temporal encoding in trace conditioning. *Animal Learning & Behavior*, **23**, 144-153.

Cole, R. P., & Miller, R. R. (1999). Conditioned excitation and conditioned inhibition acquired through backward conditioning. *Learning & Motivation*, **30**, 129-156.

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

Courville, A. C., Daw, N. D., Gordon, G. J., & Touretzky, D. S. (2003). Model uncertainty in classical conditioning. In S. Thrun, L. K. Saul, & B. Schölkopf (Eds.), *Advances in neural information processing systems* (Vol. 16, pp. 977-984). Cambridge, MA: MIT Press.

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2004). Similarity and discrimination in classical conditioning: A latent variable account. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems* (Vol. 17, pp. 313-320). Cambridge, MA: MIT Press.

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, **10**, 294-300.

Courville, A. C., & Touretzky, D. S. (2002). Modeling temporal structure in classical conditioning. In T. J. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems* (Vol. 14, pp. 3-10). Cambridge, MA: MIT Press.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience*. Cambridge, MA: MIT Press.

Dayan, P., & Long, T. (1998). Statistical models of conditioning. In M. I. Jordan, M. Kearns, & S. A. Solla (Eds.), *Advances in neural information processing systems* (Vol. 10, pp. 117-123). Cambridge, MA: MIT Press.

Dosher, B. A., & Lu, Z. L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences*, **95**, 13988-13993.

Fahle, M., & Poggio, T. (Eds.) (2002). *Perceptual learning*. Cambridge, MA: MIT Press.

Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning across tasks: A review. *Journal of Vision*, **2**, 190-203.

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

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

Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 458-467.

Fiser, J., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, **99**, 15822-15826.

Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General*, **134**, 521-537.

Furmanski, C. S., & Engel, S. A. (2000). Perceptual learning in object recognition: Object specificity and size invariance. *Vision Research*, **40**, 473-484.

Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press, Bradford Books.

Gibbon, J., Farrell, L., Locurto, C. M., Duncan, H. J., & Terrace, H. S. (1980). Partial reinforcement in autoshaping with pigeons. *Animal Learning & Behavior*, **8**, 45-59.

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron*, **31**, 681-697.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature*, **402**, 176-178.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, **49**, 585-612.

Griffiths, T. L., & Yuille, A. (2006). Technical introduction: A primer on probabilistic inference. *Trends in Cognitive Sciences*, **10**. Online supplement to Issue 7. Available at cocosci.bertesley.edu/tom/~papers/tutorial.pdf.

Hall, G., & Pearce, J. M. (1979). Latent inhibition of a CS during CS–US pairing. *Journal of Experimental Psychology: Animal Behavior Processes*, **5**, 31-42.

Haselgrove, M., Aydin, A., & Pearce, J. M. (2004). A partial reinforcement extinction effect despite equal rates of reinforcement during Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, **30**, 240-250.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. *Cognition*, **78**, B53-B64.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, **14**, 382-417.

Hunt, R., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, **130**, 658-680.

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences*, **88**, 4966-4970.

Kellman, P. J., & Arterberry, M. E. (1998). *The cradle of knowledge: Development of perception in infancy*. Cambridge, MA: MIT Press.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, **105**, 10687-10692.

Kovacs, I., Kozma, P., Fehér, A., & Benedek, G. (1999). Late maturation of visual spatial integration in humans. *Proceedings of the National Academy of Sciences*, **96**, 12204-12209.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, **5**, 831-843.

Lewicki, P., Hill, T., & Bizot, E. (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. *Cognitive Psychology*, **20**, 24-37.

Lubow, R. E. (1973). Latent inhibition. *Psychological Bulletin*, **79**, 398-407.

MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Cambridge: Cambridge University Press.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, **82**, 335-345.

Matthews, N., Liu, Z., Geesaman, B. J., & Qian, N. (1999). Perceptual learning on orientation and direction discrimination. *Vision Research*, **39**, 3692-3701.

Matzel, L. D., Held, F. P., & Miller, R. R. (1988). Information and the expression of simultaneous and backward associations: Implications for contiguity theory. *Learning & Motivation*, **19**, 317-344.

MILLER, K. D., & MACKAY, D. J. C. (1994). The role of constraints in Hebbian learning. *Neural Computation*, **6**, 100-126.

ORBÁN, G., FISER, J., ASLIN, R. N., & LENGYEL, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences*, **105**, 2745-2750.

PEARCE, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, **101**, 578-607.

PEARCE, J. M., & BOUTON, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology*, **52**, 111-139.

PEARCE, J. M., & HALL, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not unconditioned stimuli. *Psychological Review*, **87**, 532-552.

PERRUCHET, P., & PACTON, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, **10**, 233-238.

PETROV, A. A., DOSHER, B. A., & LU, Z. L. (2006). Perceptual learning without feedback in nonstationary contexts: Data and model. *Vision Research*, **46**, 3177-3197.

POGGIO, T., FAHLE, M., & EDELMAN, S. (1992). Fast perceptual-learning in visual hyperacuity. *Science*, **256**, 1018-1021.

REBER, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning & Verbal Behavior*, **6**, 855-863.

RESCORLA, R. A. (1999). Within-subjects partial reinforcement extinction effect in autoshaping. *Quarterly Journal of Experimental Psychology*, **52B**, 75-87.

RESCORLA, R. A., & WAGNER, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II*: *Current theory and research* (pp. 64-99). New York: Appelton-Century-Crofts.

SAFFRAN, J. R., ASLIN, R. N., & NEWPORT, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, **274**, 1926-1928.

SAFFRAN, J. R., NEWPORT, E. L., ASLIN, R. N., TUNICK, R. A., & BARUECCO, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, **8**, 101-105.

SCHOUPS, A., VOGELS, R., QIAN, N., & ORBÁN, G. (2001). Practicing orientation identification improves orientation coding in V1 neurons. *Nature*, **412**, 549-553.

SEITZ, A. R., & WATANABE, T. (2003). Is subliminal learning really passive? *Nature*, **422**, 36.

STADLER, M. A., & FRENSCH, P. A. (EDS.) (1997). *Handbook of implicit learning*. Thousand Oaks, CA: Sage.

SUTTON, R. S., & BARTO, A. G. (1990). Time-derivative models of Pavlovian reinforcement. In M. Gabriel & J. Moore (Eds.), *Learning and computational neuroscience: Foundations of adaptive networks* (pp. 497-537). Cambridge, MA: MIT Press.

VON DER MALSBURG, C. (1995). Binding in models of perception and brain function. *Current Opinion in Neurobiology*, **5**, 520-526.

WAGNER, A. R. (2003). Context-sensitive elemental theory. *Quarterly Journal of Experimental Psychology*, **56B**, 7-29.

WATANABE, T., NÁÑEZ, J. E., & SASAKI, Y. (2001). Perceptual learning without perception. *Nature*, **413**, 844-848.

YIN, H., BARNET, R. C., & MILLER, R. R. (1994). Second-order conditioning and Pavlovian conditioned inhibition: Operational similarities and differences. *Journal of Experimental Psychology: Animal Behavior Processes*, **20**, 419-428.