Comparing statistical learning across patterns of high and low regularity

Jordan B. Gunn

Princeton University

Author Note

This research proposal was supported by the advice and counsel of Prof. Nicholas Turk-Browne and generated to satisfy the second part of the independent work requirement for third-year students concentrating in Princeton University's department of psychology.

Abstract

Statistical learning refers to describe the ability to detect and use statistical structure. The capacity is not just a feature of human intelligence; it is ubiquitous and effective across lifespans, domains, sensory modalities and levels of consciousness. The paper outlines the power and generality of statistical learning. It proceeds to then survey the extent to which the ability varies in operation across modalities/domains and depending on prior experience, for the most part specifying some factors determining learning success. From there theoretical work on statistical learning is examined, focusing on questions concerning the unitariness, experiential sensitivity, and underlying mechanism of the ability. Building on this analysis, I outline a program of probably informative research that explores the role a sequence's statistical structure might play on the course of learning. I detail more specifically a plan for investigating how the overall contingency of patterns presented for statistical learning might impact extraction of particular transitional statistics. The particular study design enables examination of a variety of theoretical issues surrounding statistical learning so far apparently unexamined in the literature. Finally, I report the results of a short pilot of the proposed study and conclude with a discussion of potential future directions for research on statistical learning.

Keywords: statistical learning, transitional probability, summary statistics

Comparing statistical learning across patterns of high and low regularity

Suppose an agent like you or me were plopped into an environment with no detectable contingencies, where no event seemed very predictive about what has happened or might happened. The agent might try selecting actions with adaptive effect, but the effort would be broadly pointless. Nothing it perceived would reliably indicate anything about what's going on in the outside world. There would be no way to tell how actions (or inaction) might impact the world. All one could learn is that there's not much to learn.

In our world, detectable contingencies do exist, even if events still tend to be unpredictable. Across time and space and the variety of domains that specify the challenges of life, despite their essential unpredictability environments tend to be reasonably rich in patterned structure and species tend to be effective at detecting and exploiting them. The diversity of flowers is immense, but bees still exploit floral cues such as odor, shape and color to find and extract food from them (Sandoz, 2011). Similarly, by sensing a component of fish mucus, sea urchin larvae anticipate the presence of predators and bud off clones of themselves that are too small for the fish to see (Gilbert, 2012). Likewise, still when observed over short periods in the same contexts, individuals rarely behave in the exact same way on every occasion, but humans still realize that someone is likely behind a door that has just been knocked (Stamps, Briffa & Biro, 2012). This all demonstrates a straightforward point: adaptive behavior depends on reliable patterns of contingency within environments and organisms' capacity to exploit them.

Though humans and other species share this capacity, the processes by which they realize them are generally disparate. For example, many examples of species' sensitivity to environmental contingencies are rigidly determined. Bees do not learn to follow floral odor to find food; over an evolutionary timescale under a unique composite of selection pressures, development of the informationally contingent behavioral pattern became a nearly universal dictate of the species' genetic code (Hansson & Stensmyr, 2011). Humans, on the other hand, exhibit from infancy a capacity for *statistical learning,* a more-or-less domain general ability to extract structure from patterned input over very short timescales (Lew-Williams & Saffran, 2012). Working quickly across a broad variety of tasks and stimulus modalities to detect a diversity of statistics about environments, statistical learning constitutes a powerful feature of human intelligence inaccessible to most other species.

This paper can be considered a comprehensive review of the research literature on statistical learning and presentation of a novel study idea for advancing theoretical work in the field. I begin with a broad positive description of our capacity for statistical learning, delineating the power, efficiency and generality of the process as demonstrated across a diverse body of experimental work. Next, I survey the extent to which statistical learning varies in operation across modalities and domains, as well as depending on prior experience, for the most part specifying some factors determining learning success. From there theoretical work on statistical learning is examined, focusing on questions concerning the unitariness, experiential sensitivity, and underlying mechanism of the ability. I outline a program of probably informative research that explores the role a sequence's statistical structure might play on the course of learning. I detail more specifically a plan for investigating how the overall contingency of patterns presented for statistical learning might impact extraction of particular patterns, just as I have in this introduction comparing statistical learning between highly ordered and highly disordered contexts. Finally, I report the results of a short pilot of the proposed study and conclude with a discussion of potential future directions for research on statistical learning.

The Power and Generality of Statistical Learning

Statistical learning is the most common term used to describe the ability to detect and use statistical structure (Alvarez, 2011). It seems appropriate to begin with an accounting of the features of the capacity because doing so seems to achieve two functions. First, it helps illustrate why and how much statistical learning matters as an object of study. Statistical learning is not just necessary; it is ubiquitous and adaptive across lifespans, domains, sensory modalities and levels of consciousness. Beyond that, though, an accounting of statistical learning's features – its apparent power and breadth of contribution to cognition – makes later circumscription of the ability's limitations clearer. Evidence outlining the power and generality of statistical learning are described here:

Statistical learning works from infancy

Some of the earliest work on statistical learning demonstrated its operation early in life. One experiment by Saffran, Aslin and Newport (1996), for example, examined its role in the segmentation of words from fluent speech, an important marker of language acquisition. Eight-month-infants listened to a continuous, monotonic speech stream consisting of four three-syllable nonsense words repeated in random order. Every cue to word boundaries were removed, leaving only the transitional probabilities between syllabus pairs. Despite this, infants indicated that they were familiar with the unique statistical structure of this stream by not listening as long to words rooted from it as they would to novel non-words. Kirkham, Slemmer & Johnson (2002) in fact found evidence of statistical learning from an even younger age – 2-month-old infants through a similar research paradigm demonstrated discrimination between familiar and novel patterns of shapes.

Statistical learning occurs unconsciously, automatically and quickly

The infant studies just described do not merely illustrate that statistical learning is an ability available to humans from very early age; they show that statistical learning occurs automatically and apparently effortlessly (without exercising cognitive control), with only the presentation of a stimulus pattern necessary to trigger the process. Furthermore, this process of statistical learning proceeds rapidly; the words that babies became familiarized with in the Saffran et al (1996) study were presented within an only 2-minute-long stream. Studies involving adults demonstrate a similar pattern. For example, statistical patterns shown to participants while completing an unrelated task still caused bidirectional associative shaping in brains indicative of rapid and incidental statistical learning. Triesman & Chong (2004) found that statistical learning of mean size is not disrupted when cover tasks require either distributed or global attention to an artificial environment, though focusing attention on specific objects within an environment could disrupt learning. All of this suggests that statistical learning is a highly efficient process requiring no especial effort or control of behavior.

Statistical learning enables apprehension of a broad variety of statistics

The variety of statistics that individuals are apparently able to extract from their environments is also noteworthy. As already noted, Saffran, Aslin and Newport (1996) along with other researchers demonstrated that infants can extract transitional statistics, probabilities that a given unit will follow another specified unit in a sequence, from a stream of syllables. These can be thought of as examples of a broader class of statistics called conditional statistics, the predictive relationship between two events X and Y. A related sensitivity that humans exhibit is to cue-based statistics – relationships between perceptible attributes and attributes not directly perceptible, such as to emotion (Theissen, Kronstein, & Hufnagle, 2013). Individuals compute comparatively distributional statistics readily as well. For example, they evaluate the randomness of sequences in the same (somewhat biased) way across a variety of contexts, feature dimensions, sensory modalities, speed and manner of presentation (Yu, Gunn, Osherson & Zhao, Unpublished). Chong & Triesman (2004) found that judgment of the mean size of a set of circles are nearly as accurate as judgments of the size of single circles presented alone. Not only can humans judge the average emotion expressed by a crowd; they are also sensitive to the variance or heterogeneity of facial expressions within the same crowds (Haberman, Lee & Whitney, 2015). The list of statistics that humans seem to be able to extract through statistical learning is vast; it can only be partially delineated here.

Statistical learning enables performance in a variety of tasks, domains, and modalities.

The variety of domains in which these described statistics can be extracted is similarly enormous. Individuals indeed can judge the mean size of circles efficiently as recounted earlier (Chong & Triesman, 2004), but they can also extract the average emotional expression, gender, identity, gaze direction, and ethnicity of members in a crowd (Haberman et al, 2015). In the same way, while individuals can from infancy extract transitional statistics from a sequence of heard verbal syllables (Saffran et al, 1996), sequence statistics can be extracted without much regard to the sensory modality or process in which said sequences are presented. Learners have displayed sensitivity to sequential conditional statistics in tactile, visual and tonal stimuli (Thiessen, Kronstein, & Hufnagle, 2013). And along with helping to segment continuous auditory input (Saffran et al, 1996), evidence for statistical learning has been found to facilitate visual search, contextual cuing, visuomotor learning, conditioning and generally any predictive behavior (Frost, Armstrong, Siegelman & Christiansen, 2014).

Clearly, statistical learning is a powerful, broadly functional feature of human cognition.

Factors Constraining Statistical Learning

Despite this huge base of evidence outlining the adaptive use of statistical learning, humans are of course not perfect learners, and the literature outlining constraints on statistical learning is around as substantial as the literature delineating its breadth and efficiency. In particular, stimulus presentation, sensory modality, task domain, extracted statistic, and past experience have all been identified as profoundly impactful on the course and effectiveness of statistical learning.

Statistical learning across modalities and stimuli

Despite the apparent domain generality of statistical learning, evidence has emerged that the process is subject to modality and stimulus-specific constraints (Frost et al, 2015). In a review of evidence for transfer in artificial grammar learning, for example, Redington & Chater (1996) found only small and ambiguously meaningful magnitudes of artificial grammar learning effects upon modality transfer. This separation of statistical learning between modalities may in fact have adaptive consequences: while transfer of learning between modalities may be limited, simultaneous learning of two sources of statistical structure can occur without mutual interference if implemented in distinct modalities (Conway & Christiansen, 2006). Other qualitative differences emerge across modalities in patterns of statistical learning, too. For example, compared to tactile and visual statistical learning, auditory statistical learning is easier and better for the final part of input sequences than for earlier parts (Conway & Christiansen, 2005). Even within modalities, no statistical learning is transferred upon alteration of the stimuli by which a statistical structure is presented (Conway & Christiansen, 2006). Indeed, evidence suggests that as more time is taken to observe a structured sequence, knowledge of its statistical structure becomes more stimulus-specific rather than abstracted towards surface-independent representations (Johansson, 2009). These facts complicate the notion that statistical learning is a domain-general process. Though statistical learning works *within* many different modalities, in the strictest sense of the phrasing, it does not work much *across* those modalities. They recommend that a more precise conception of statistical learning must emerge beyond that of a unitary system indifferent to the manner by which statistics are presented to the learner.

Statistical learning across tasks/statistics/domains

In the same way, evidence has emerged that statistical learning operates differently depending on the statistics extracted to fulfill the demands of a task. For example, something about the process by which humans judge the randomness of sequences and other stimuli demonstrates a consistent bias in that form of statistical learning which is not evident in the way humans extract other statistics. In the case of randomness perception, humans generally exhibit negative recency, “expectations that a streak of events will end” (Oskarsson et al, 2009). They are more likely to judge sequences with short event streaks than truly random sequences as random. However, in certain domains, this pattern reverses: in domains involving wins and losses, especially in gambling scenarios, humans instead exhibit positive recency, expecting streaks of events to continue. The fact that evidence for similar patterns of bias have not emerged in study of other sorts of statistical learning suggests that the computation of different statistics might involve different processes, complicating the notion that some unitary system does the work, or even that “statistical learning” refers to a single topic at all.

More broadly articulated examples of this conflict exist. Aslin & Newport (2016) raise the possibility that learning about elements that have presented during exposure is an achievement distinct from that of learning rules that can be applied to novel elements and novel combinations. Along a similar vein, Thiessen et al (2013) draws a distinction between the extraction of discrete representations such as that of particular conditional statistics or words, and the extraction of distributional statistics that characterize the prototypical characteristics of an element set. Rooted in these distinctions is not a particular set of experimental results so much as a considered analysis of differences between the tasks learners perform well on. Generalizing learned statistics to apply to future stimuli or experiences seems to be a step that can *only* occur after and distinctly from actual learning of those statistics, according to Aslin & Newport (2016). Similarly, computation of the broad distributional statistics of a pattern seems to *require* integrating previously extracted low-level discrete statistics, according to Thiessen et al (2013). Despite the intuition and logic supporting these divisions, further considerations are necessary to determine whether they aptly characterize distinct processes within minds.

Statistical learning across experience

Along with previously described factors, the course of statistical learning also seems to be profoundly sensitive to past experience. Nine- and 10-month-old infants exposed to sequences of either disyllabic or trisyllabic words, for example, only successfully segment words within a subsequently presented sequence if the sequence is equally disyllabic or trisyllabic, suggesting that prior experiences of statistical structure equip infants with expectations about future statistical structure (Lew-Williams & Saffran, 2012). Another experiment similarly presented adults with two sets of patterns – either first an unordered ‘pattern’ and then an ordered pattern, or first an ordered pattern and then an unordered pattern (Junge, Scholl, & Chun, 2007). Only participants in the order-first condition displayed learning of the ordered pattern later on, similarly suggesting that prior experiences of statistical structure impact expectations about future experiences about statistical structure. The work goes further than that, though – it suggests that a reverse effect does not exist, at least as strongly: experiencing a novel statistical structure does not substantially impact learning of the statistics extracted immediately before the new experience. Interestingly, though, a study by Gebhart, Aslin, & Newport (2009) replicated these findings but discovered that explicitly indicating the presence of two statistical structures or extending experience to the second statistical structure can eliminate these effects. The impact of experience on statistical learning may therefore be dependent on the extent to which prior experience can be judged to be relevant.

Theoretical Issues

Do different processes underlie statistical learning?

The issue of whether statistical learning is one process actually turns on two issue. First, it is unclear if a single process performs statistical learning between stimuli or sensory modalities. Second, it is unclear if a single process can compute the broad variety of statistics that characterize statistical learning.

Aslin & Newport (2016) raise the possibility that learning about elements that have presented during exposure is an achievement distinct from that of learning rules that can be applied to novel elements and novel combinations. However, it is possible that additional rules do not have to be represented in order for learned structure to impact later experience. For example, extant grouping principles like temporal proximity, perceptual similarity and shared context might cause the individual to observe stimuli as exemplary of past experience. Thiesson et al (2013), however, assert that the extraction of consistent clusters in the input cannot by itself enable the ability to generalize from prior experience. Furthermore, they assert that additional mechanisms are necessary to account for our sensitivity to distributional statistics like average size and variance. They thus account for statistical learning in terms of two major processes. First, individuals are thought to *extract* clusters of the input with processes and represent them discretely. Second, individuals are thought to *integrate* across these clusters to represent the central tendency of exemplars stored in memory and benefit from their similarity and distribution, which then guides subsequent extraction. This distinction is rooted in a much deeper computational analysis of the tasks involved in statistical learning than that achieved by Aslin & Newport (2016). Because it casts statistical learning as dependent on two mutually dependent processes, too, it doesn't quite defeat the idea that statistical learning operates as a single system.

The considerable similarities in our statistical learning capacities across modalities inclines one to presume that a single, stimulus-independent process carries all the learning out, but the limited transfer of learning between the modalities constitutes a serious challenge to that position (Frost et al, 2015). However, one line of argument might help disarm that particular challenge: failure to transfer learning between sensory modalities might be a consequence of the same processes that govern how we determine if extracted statistics are likely to apply to new experiences. Since cross-sensory stimuli are so dissimilar, limited transfer of learning between sound and touch might be just an extreme example of the principles that prevent learning transfer between two same-modality structures separated only by a 30s pause (as demonstrated by Gebhart et al, 2009). Other challenges to the unitary model of statistical learning are more substantive. For example, Frost et al (2015) report (but have no yet published) evidence suggesting that there is no significant correlation between individuals' performance on statistical learning tasks between sensory modalities as might be predicted if statistical learning were a unitary process. It's still unclear, then, whether statistical learning is a mostly unitary process or a set of similar computational processes implemented independently within different sense-specific systems.

Why does experience impact learning the way it does?

Each piece of literature illustrating the impact of prior experience on statistical learning reliably proposes a relatively simple mechanism for the process. First, individuals experience and extract from some set of stimuli a particular statistical structure. Then, because these experiences have set hypotheses or expectations for future experience, they treat further experiences/stimuli as having the same (or at least similar) statistical structure, and only stop if extended experience makes evident that this treatment is inappropriate (Lew-Williams & Saffran, 2012; Gebhart et al, 2009; Junge et al, 2007). However, it is not obvious why it should take longer for learners to extract statistics after having extracted other statistics than after having extracted no statistics at all; the existence of these expectations alone cannot explain the impact that past experience of structure has on future learning of different structure. Gebhart et al (2009) argue that our cognitive architectures make an organizational trade-off between the inefficiency involved in waiting for an input to be extracted and that in having to recover from errors driven by biasing learning in favor of early input. Here, the primacy effect is a deliberate heuristic that enables reaction to ordered stimuli faster than the time it takes for an entire body of stimuli to be extracted. The fact that this trade-off is moot when explicit cues for distinction between statistical structures are evident may limit the harmful effects of the bias to fewer contexts.

Proposal

The Basic Design

As a novel direction for research on statistical learning, I propose examining the way statistical properties of sequences can impact the way we pick up on particular statistical clusters (i.e., transitional relationships between units X and Y) within said sequences. There's a straightforward experimental logic in doing this sort of research, independent of particular details that might define a specific experimental design:

The “basic design” involves two groups of participants (Group 1 and Group 2), clearly defining a between-subjects design. Membership within groups is randomly assigned. Both groups are somehow shown statistically defined sequences and later somehow tested on their apprehension of patterns within these sequences. Between subject groups, units within a sequence are presented in the same way, along the same medium, using the same sorts of stimuli; similarly, their apprehension of patterns within these sequences are measured in the same way. Instructions given to participants to appropriately guide their behavior are likewise identical across groups.

The only thing different between these groups is the statistical distribution of units within the sequences the groups are presented. Even then, most features of the distribution are chosen to be held constant between subject groups. The independent variable in this experiment is the facet of the sequences’ units’ statistical distribution that is systematically manipulated between groups. The dependent variable is not necessarily merely observers’ overall performance on some measurement of their apprehension of the sequence’s underlying statistical structure; how people learn about patterns controlled between conditions should be understood as the *main* dependent variable in the experiment. The result is a rather powerful general strategy to answer an assortment of questions about how the overall statistical structure of sequences affects people's ability to pick up on particular patterns within them. Despite the straightforwardness and power of this paradigm, no research in the field has applied it yet, instead focusing on adjacent but still interesting questions about statistical learning. The remainder of this paper is devoted to detailing an example of a potential research study based on this basic design.

Examining Overall Contingency

It is asked how manipulating the overall contingency of patterns within a sequence might impact statistical learning of specific patterns within the same sequence. There may be several ways an experiment delineated by the basic design just described might seek to answer this question, but I select on that defines the statistical structure sequences almost totally in terms of transitional probability relations between clusters of units:

Say the possible units within sequences can be identified as A, B, C, D, E, F, G and H. These units are defined as such wholly by the way they are differentiated during stimulus construction and presentation; within sequences they are repeated and organized according to a set of statistical rules. According to the strategy of defining the sequence in terms of transitional probability relations, each unit is paired off into statistical clusters – AB, CD, EF, and GH. The contingency of these clusters defines the reliability of transitional relationships between clustered units across a sequence. Highly contingent clusters between hypothetical units X and Y, for example, might feature a transitional probability between the two of much more than .5.

When I speak of overall contingency, I'm referring to the average contingency of every cluster defined within a sequence. When a sequence's coherence is very high, its units are organized in a very orderly fashion; everything about it is profoundly predictable. At the other extreme, less contingent sequences are organized almost randomly. The sequence for the first group (Group 1) might establish a transitional probability of .5 between A and B, and a transitional probability of .7 for clusters CD, EF and GH, making its overall contingency relatively high. In other words, 70% of the time, when C occurs, D will occur next; otherwise some other letter may occur next instead –and so forth. Group 2, on the other hand, might establish a transitional probability of .5 for AB, but a transitional probability of .3 for EF and GH. Here, the contingency of clusters AB is the same between conditions (.5), and weaker for CD, EF and GH in one condition (.3) but stronger in the other (.7). This way, after showing participants patterns based on either this first or second set of rules, a scientist can measure and analyze how the difference of overall contingency between conditions impacted participants' learning of the constant cluster AB.

In addition to these constraints, the frequency of each unit within the sequence is required to be equal to one another, ensuring that differences in experience with each unit are not a possible explanation for differences in measured amount of learning. Importantly, sequences are not just generated probabilistically according to these rules! Instead, specified quantities of units are arranged and rearranged stochastically until they reflect the rules enforced (for example, until B is followed by A exactly 50% of the time). For this reason, the probability distribution of units that may follow A in a sequence if B does *not* follow A is not uniform. When frequency is controlled, for example, pairing CD with the maximum contingency 1.0 means not only that in a produced sequence D follows C 100% of the time, but also that D will never occur after some other letter; even when contingency is lower, like .7, there are still less Ds to spare that might follow another letter (like, for instance, A). This pattern becomes relatively inconsequential as statistical learning carries out, and ensures that unit frequency cannot explain learning while at the same time enforcing the controlled and manipulated transitional statistics between conditions.

Possible outcomes

Why might a scientist be interested in this particular question, though? That is, why should a research program concerning the way sequence features impact statistical learning begin with a focus on overall contingency? The most straightforward answer to that question is simply that results would inform theoretical work on statistical.

We can distinguish at least three broad possible results of the experiment, each of which has some intuitive and empirical support:

Statistical learning of moderately contingent clusters might be comparatively enhanced within highly contingent sequences. We know that humans can compute distributional statistics representing the overall orderliness of sequences (Yu et al, 2015). Perhaps this statistic impacts sensitivity to transitional statistics through some sort of confirmation bias, at first preparing observers to expect or not expect to find clusters within sequences depending on how overall contingency impacts initial impressions about said sequences. When initially appraised orderliness is high, confirmation bias might increase sensitivity to particular clusters; when it is instead low, confirmation bias might decrease sensitivity to clusters. Since we know that statistical learning tends to exhibit a primacy effect when updating representations (Gebhart et al, 2009; Junge et al, 2007), this account isn't implausible.

Another set of potential accounts, rooted in the general features of chunking models of word segmentation (e.g., Perruchet & Vinter, 1998), emphasize the impact of cluster contingency on the degree to which concurrently learned clusters interfere with apprehension of moderately contingent clusters. One version of this relies on the intuition that highly contingent clusters will be extracted more quickly than less contingent clusters. Once a cluster is extracted, a number of candidate clusters are immediately ruled out, potentially facilitating learning of other clusters.

For example, once an observer learns that CD in our design constitutes a cluster, they might rule out candidate clusters like AD and learn actual cluster AB more quickly. This learning pattern is less likely to arise in the low contingency condition. A closely related account might assert that the high contingency of other clusters curtails learning interference *before* any clusters are confidently extracted.

Statistical learning of moderately contingent clusters might be comparatively diminished within highly contingent sequences. It is possible that participants' computed distributional statistics representing a sequence's overall orderliness might impact learning of moderately contingent clusters in a different way from that described earlier. Instead of creating a confirmation bias, the representation might modulate one's "standard" or "threshold" for positive detection of a contingency between units across a sequence. A highly contingent sequence might modulate this threshold upward, causing moderately contingent relations to seem random in comparison to either one's perception of the sequence's overall contingency or the more strongly contingent clusters they occur around it. On the other hand, less contingent sequences might facilitate increased sensitivity to moderately contingent clusters occurring in a comparatively disorderly context through some sort of pareidolia effect (Voss, Federmeier & Paller, 2012). Yet another possibility is that relatively contingent statistical clusters perceptually "pop out" from less contingent sequences in the same manner that a red circle may perceptually pop-out on a page of blue circles (Treisman, 1985).

A final potential account might rely on the capacity limitations of short-term memory (Cowan, 2008). According to this reasoning, the amount of time and effort required to hold an additional cluster in working memory might increase with the number of clusters already being held in memory. If this is true, learning some clusters early may interfere with and increase the amount of time required to learn other clusters. In a sequence organized according to the high overall contingency condition, highly contingent clusters may be more obvious and learned first, making it harder to learn moderately contingent clusters; in the low overall contingency condition, moderately contingent clusters might be learned first and no interference would occur.

Statistical learning of moderately contingent clusters might be invariant with respect to the overall contingency of patterns within a sequence. It could simply be that neither the ensemble perception of orderliness nor concurrent cluster learning enhances or diminishes statistical learning of moderately contingent clusters. Some theoretical work already exists asserting the independence in the manner by which different extracted statistical structures are represented cognitively (Gebhart et al, 2009); it is possible that this representational independence or other facet of statistical learning limits interaction between extracted statistics. This null result may thus have powerful implications about the nature of statistical learning. Another possible account of this result exists as well, however: it could be that humans are too insensitive to differences in strengths of transitional probabilities of .4 or .2 for the design as described to have a measurable effect on behavior. Research elsewhere though suggests that learners are relatively sensitive to statistical coherence in stimuli streams, making the accuracy of this account less credible (Thiessen & Saffran, 2003).

The breadth of possible accounts for different outcomes of the proposed experiment evinces the theoretical import of this particular experimental design. Whatever the outcome of the proposed experiment, many apparently intuitive theoretical conceptions of statistical learning will be substantiated or discredited.

Method

Here the details of the proposed experiment are specified fully. The sequence could be presented and statistical learning measured in a variety of different ways, but the chosen method is similar to that applied in other demonstrations of statistical learning, offers no cues signaling statistical boundaries,

Participants

Following the lead of other research in the field, between 20 and 25 participants per condition is expected be enough to produce a study of reasonable power that demonstrates the effect of statistical learning and illustrates the impact of the independent variable.

Apparatus

The experiment was programmed in Matlab using the Psychophysics Toolbox extensions (Brainard, 1997). It will be conducted on a Windows desktop computer with 17-inch screen and subjects will view the display without restraint from a nearby seat in a sound-controlled room.

Sequencing

Principles of organization. Two constraints determine how sequences are generated. First, the frequency of units within each sequence is equated. For example, unit A will occur as often as unit B and C and so forth. Second, certain transitional statistics between pairs of units are enforced. In both conditions, B follows A 50% of the time in sequences. The contingency of other clusters (CD, EF, GH) vary between conditions. They are .3 in the low overall contingency condition, and .7 in the high overall contingency condition. **Table 1** summarizes these principles. For each of 7 blocks, a sequence 96 instances of units long is produced.

Algorithm. In order to produce sequences following these constraints, first the specified frequency of each unit was generated. In a sequence of 96 instances, units were repeated 12 times each. Next, units were clustered according to the transitional statistics to be enforced. For example, in both conditions, B follows A 50% of the time; therefore half of the Bs and half of the As were combined to make 6 AB clusters, leaving 6 independent A clusters and 6 independent B clusters. Once sorted into clusters this way according to transitional statistics, the pool of clusters were shuffled into a sequence. Finally, the transitional and frequency statistics of the sequence were tested to confirm enforcement of constraints.

Procedure

Presentation. The sequence was presented as a temporal series of distinctly colored squares presented at the center of a grey screen (**Figure 1**). Units of the sequence (A, B, C, D…) were represented as the colors of these squares. Possible colors were red, green, blue, cyan, yellow, magenta, brown and black and were randomly assigned to units at the start of each experiment; this assignment was constant throughout blocks. Each colored square was shown in the screen's center for 1000ms; in between each square presentation, a fixation point was shown at the screen's center for 500ms. Within each of 7 blocks, 96 instances of units would be presented this way; the frequency of each unit was held constant so that 12 instances of each unit occurred in every block.

Testing. A set number of times within each block, instead of the next colored square from the sequence, only the square's outline and a color wheel will be shown (**Figure 2**). Participants are instructed to move the mouse to the color wheel entry that matches their best guess of the square's color based on their "experience so far". Upon movement of the mouse (represented a crosshair) to a color wheel entry, the response and reaction time is stored and sequence presentation continues after an interstimulus interval of 500ms and presentation of a fixation point for another 500ms. Prediction of the most likely subsequent square is tested for each possible unit (A, B, C…) twice, resulting in 16 tests per block and 16 x 7 = 112 tests throughout the experiment. Tests were arranged randomly, but interspersion was enforced such that three letters of the sequence were shown before an additional test could occur.

End of Block. At the end of each block, participants will be allowed to take a short break of unspecified length. Participants will be informed of the proportion of tests in which they accurately guessed the *most likely* unit to occur given the unit that had occurred most recently in the sequence. Finally, participants will again be instructed to move the mouse to the color wheel entry that matches their best guess of the square's color depending on their experience of the sequence so far.

Analysis

ANOVA and related statistical tests would be applied to compare and examine the distribution of prediction accuracy across subjects, conditions, particular clusters, experience (as indicated by the block in which a prediction is made), and reaction times. To ensure that the colors used to designate units does not impact learning, learning rate of equivalently contingent clusters represented by distinct sets of colors will be compared. Along with learning rate, response biases will also be analyzed – are participants inclined to select particular colors, for example? Similarly, relationships between reaction times and other variables such as stimuli, block, and prediction accuracy will be compared in order to confirm that decision times relate only with experiment parameters expected to relate with successful statistic extraction (experience and prediction accuracy). An additional insanity check would confirm that highly contingent clusters (i.e., with .7 contingency) are extracted more quickly and more successfully than less contingency clusters (.5 or .3 contingency) and similar results that are almost taken for granted in the design of the experiment.

The comparison focal to the experiment, however, will be of prediction accuracy of the controlled .5-contingency AB cluster between the two conditions. Prediction accuracy that B follows A might be significantly different between the courses of the two experimental conditions overall, or within particular blocks. Either finding, combined with the direction of found difference or the null possibility that no difference between the two conditions on this measure is detected, will define the main conclusion of the proposed experiment, as outlined earlier in this paper.

Pilot

In order to work out and test details of the experiment design, a small pilot involving 10 subjects was performed implementing most but not all of the previously described procedures.

Participants

10 naïve participants (6 women; median age 21 years) were recruited by convenience to participate in the pilot. 5 were assigned to the high contingency condition while the other 5 were assigned to the low contingency condition.

Apparatus

The pilot was programmed in Matlab using the Psychophysics Toolbox extensions (Brainard, 1997). It was conducted on a Windows laptop computer with 14-inch screen and subjects viewed the display without restraint from a nearby seat in a quiet location.

Presentation

Stimuli were organized and presented as designed in the experiment, with the modification that only 5 blocks were complicated (in order to save participants' volunteered time).

Results and Analysis

Due to the small sample size and limitations of time, analysis was cursory. Significance tests are almost certain to fail to reject the null hypothesis, so emphasis is placed on identifying trends and Mean reaction in the low contingency condition was 2.678s in the first block and 2.271s in the final block; similarly in the high contingency condition, mean reaction time in the first block was 2.244 and in the final block 2.095. Overall, between reaction time and trial number, a correlation of -.153 was found in the low contingency condition and -.065 in the high contingency condition. Reaction time thus behaves as expected in the experiment – falling gradually through experience, but generally being higher when the task is harder (as it should be in the low contingency condition), suggesting that confidence with the task will explain any detected differences in reaction time.

What about performance? Accurate prediction rate over all units in the high contingency condition is .0625 in the first block and rises to .2533 in the final block on average. In the low contingency condition, rate is .1 in the first block and .1875 in the fifth block on average. Focusing on particular clusters, low contingency (.3) clusters were accurately predicted at a rate of .3975, with a rate of .5 in the final block. High contingency clusters were accurately predicted at a rate of .6055 overall, with a rate of .863 in the final block. As expected, learning occurs, clusters in the low contingency condition are extracted less fully by the end of the experiment than clusters of the highly contingent condition. A full experiment will be able to confirm whether the difference in learning is significant or not, particularly when the experiment is extended over a greater number of blocks, providing a greater opportunity for learning.

More specifically, we examine how learning of the controlled AB cluster differs between conditions. Mean performance for the AB cluster was .489 across the experiment in the low contingency cluster, and .5 in the high contingency cluster. Accuracy during the first half of tests in the low contingency condition was .26, and .24 in the high condition. By the final block of the experiment, however, performance in the high contingency condition was .75 and in the low contingency condition was .708. The results thus seem to suggest that the null hypothesis is correct, that no difference in extraction of AB will occur between conditions.

Discussion and Conclusion

Our ability to extract and exploit statistical structure is fundamental to our intelligence, ubiquitous and adaptive across lifespans, domains, sensory modalities and levels of consciousness. However, stimulus presentation, sensory modality, task domain, extracted statistic, and past experience have all been identified as profoundly impactful on the course and effectiveness of statistical learning. These diametrical trends in the literature permit one to draw profoundly diverse conclusions about the nature of the process. Is statistical learning a unitary, domain-general process that processes information independently of the way it is presented, or is it a complex of processes defined by modality and sort of statistic extracted that interact with each other only to a limited extent? Discourse within the literature sits in the middle, but for now it seems that statistical learning is more like a domain-general process than many, many independent ones.

Unlike most other experimental work on statistical learning, which tends to focus on how statistical structure is presented or which statistical structures are presented beforehand, my proposed study enables analysis of how the features of a statistical structure might impact extraction of it. Beyond simply confirming that the structure more predictably organized sequences can be extracted more quickly than that of less predictably organized sequences, the proposed experiment enables examination of how statistical learning operates within *contexts* of relative order and disorder. In a context defined by many patterns, might the characteristics of some of those patterns influence the learning of others?

In answering that question, a broad variety of theoretical issues concerning statistical learning can begin to be illuminated. Most significantly, the extent to which representations of particular statistics are independent or interact within minds can be measured in this way.

Differences observed through the implementation of my experiment may be driven by extraction of particular clusters or of broad distributional statistics or both, but in either case the proposed design enables measurement of the extent of this interaction. Furthermore, the proposed design may enable extension or complication of work analyzing how *previous* experience impacts statistical learning to help draw conclusions about how *ongoing* experience might impact statistical learning. Since varying the overall contingency of clusters between conditions in my proposed experiment might also vary the *order* in which clusters are generally extracted, the experiment may realize through the presentation of a single statistical structure processes that so far have only been identified through the presentation of alternating statistical structures.

The breadth of possible accounts for different outcomes of the proposed experiment do not just highlight the starkly contrasting models of learning that the experiment might adjudicate between. They also circumscribe different models of learning that the current design cannot adjudicate between alone, signaling future directions for research depending on the experiment's results. A broad distinction emerges in possible result explanations between *sensitivity* to the distribution statistics of the presented sequences and sensitivity to the particular clusters that determine the overall contingency of the presented sequences. By one logic, participants' perception of the general orderliness of sequences impacts sensitivity to moderately contingent clusters; by another, being concurrently exposed to differentially contingent clusters as learning occurs impacts sensitivity to moderately contingent clusters in a more distributed, bottom-up fashion.

In order to investigate these differing accounts, an experimenter might vary the contingency of particular clusters within a stimulus sequence while controlling the *overall* contingency of the sequence's clusters (**Table 2)**. For example, consider two conditions defined by 10 unit sequences. Instead of just making possible A, B, C, D, E, F, G, H, also possible are I and J. Clusters are defined according to transitional probabilities similarly to the previous design: AB, CD, EF, GH and IJ are all assigned a specific amount of contingency. Under both conditions, AB's contingency is once again controlled at .5. However, in the first condition, CD, EF, GH, and IJ are all *also* set at .5 contingency. In the second condition, though, CD and EF are set at .3 contingency but GH and IJ are set at .7. In a sense, both sequences are equally orderly/random: in both conditions, predictions of the next unit given the current unit based on knowledge of the clusters will be accurate 50% of the time. However, the way predictability is distributed within the sequence is modified: in one condition, predictability is uniform and in the other, it varies from .3 to .7. In this way, overall contingency is controlled, but the contingency of particular clusters within the sequence is varied. With learning of the constant cluster AB measured, the design might adjudicate between the two possible accounts of the results of our first experiment. Supposing that participants accurately perceive overall order/randomness between these sequences as equivalent (a finding can that can itself be confirmed or rejected through a very similar design), learning of AB between the two conditions will either be distinguishable or not. If learning *is* different, then an extracted distribution statistic of order/randomness cannot explain the results of our currently proposed experiment. A model must emphasize the role of concurrent extraction of instances of transitional statistics in constraining or enhancing statistical learning. Of course, if learning *isn't* different, then many different possible explanations of the results remain live, exposing a flaw in this particular design that perhaps careful consideration of the results of the currently proposed experiment might help avoid.

Still, the research design proposed here does a lot of work on its own to rule out a range of possible models of overall contingency as a parameter of statistical learning. It can serve as the beginning of a line of research on statistical learning that illuminates representations and mechanisms guiding the process in a way not yet been pursued elsewhere in the literature.

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. Trends in Cognitive Sciences, 15(3), 122-131. doi:10.1016/j.tics.2011.01.003

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. Cognition, 106(3), 1382-1407. doi:10.1016/j.cognition.2007.07.005

Conway, C. M., & Christiansen, M. H. (2005). Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(1), 24-39. doi:10.1037/0278-7393.31.1.24

Conway, C. M., & Christiansen, M. H. (2006). Statistical Learning Within and Between Modalities: Pitting Abstract Against Stimulus-Specific Representations. Psychological Science, 17(10), 905-912. doi:10.1111/j.1467-9280.2006.01801.x

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? Progress in Brain Research Essence of Memory, 323-338. doi:10.1016/s0079-6123(07)00020-9

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. Trends in Cognitive Sciences, 19(3), 117-125. doi:10.1016/j.tics.2014.12.010

Gilbert, S. F. (2012). Ecological developmental biology: Environmental signals for normal animal development. Evolution & Development, 14(1), 20-28. doi:10.1111/j.1525-142x.2011.00519.x

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing Structures in Midstream: Learning Along the Statistical Garden Path. Cognitive Science, 33(6), 1087-1116. doi:10.1111/j.1551-6709.2009.01041.x

Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. Journal of Vision, 15(4), 16. doi:10.1167/15.4.16

Hansson, B., & Stensmyr, M. (2011). Evolution of Insect Olfaction. Neuron, 72(5), 698-711. doi:10.1016/j.neuron.2011.11.003

Johansson, T. (2009). Strengthening the Case for Stimulus-Specificity in Artificial Grammar Learning. Experimental Psychology, 56(3), 188-197. doi:10.1027/1618-3169.56.3.188

Jungé, J. A., Scholl, B. J., & Chun, M. M. (2007). How is spatial context learning integrated over signal versus noise? A primacy effect in contextual cueing. Visual Cognition, 15(1), 1-11. doi:10.1080/13506280600859706

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. Cognition, 83(2). doi:10.1016/s0010-0277(02)00004-5

Lew-Williams, C., & Saffran, J. R. (2012). All words are not created equal: Expectations about word length guide infant statistical learning. Cognition, 122(2), 241-246. doi:10.1016/j.cognition.2011.10.007

Matthias, B. (2012). How Sensitive Is the Human Visual System to the Local Statistics of Natural Images? Front. Comput. Neurosci. Frontiers in Computational Neuroscience, 6. doi:10.3389/conf.fncom.2012.55.00053

Oskarsson, A. T., Boven, L. V., Mcclelland, G. H., & Hastie, R. (2009). What's next? Judging sequences of binary events. Psychological Bulletin, 135(2), 262-285. doi:10.1037/a0014821

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. Journal of Experimental Psychology: General, 125(2), 123-138. doi:10.1037/0096-3445.125.2.123

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. Science, 274(5294), 1926-1928. doi:10.1126/science.274.5294.1926

Sandoz, J. C. (2011). Behavioral and Neurophysiological Study of Olfactory Perception and Learning in Honeybees. Front. Syst. Neurosci. Frontiers in Systems Neuroscience, 5. doi:10.3389/fnsys.2011.00098

Schnee, J. E. (1977). Predicting the unpredictable: The impact of meterological satellites on weather forecasting. Technological Forecasting and Social Change, 10(3), 299-307. doi:10.1016/0040-1625(77)90026-9

Stamps, J. A., Briffa, M., & Biro, P. A. (2012). Unpredictable animals: Individual differences in intraindividual variability (IIV). Animal Behaviour, 83(6), 1325-1334. doi:10.1016/j.anbehav.2012.02.017

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. Psychological Bulletin, 139(4), 792-814. doi:10.1037/a0030801

Voss, J. L., Federmeier, K. D., & Paller, K. A. (2011). The Potato Chip Really Does Look Like Elvis! Neural Hallmarks of Conceptual Processing Associated with Finding Novel Shapes

Subjectively Meaningful. Cerebral Cortex, 22(10), 2354-2364. doi:10.1093/cercor/bhr315

Tables

Table 1

Contingencies of Clusters between Conditions in Proposed Experiment

	AB	CD	EF	GH	*Average*
High Contingency Condition	.5	.7	.7	.7	.65
Low Contingency Condition	.5	.3	.3	.3	.35

Note: “Contingency” is here defined as the likelihood that the second unit in a cluster will follow the first within a sequence. For example, in both conditions, B will follow 50% of the occurrences of unit A.

Table 2

Contingencies of Clusters in Proposed Follow-Up to Proposed Experiment

	AB	CD	EF	GH	IJ	*Average*
High Contingency Condition	.5	.5	.5	.5	.5	.5
Low Contingency Condition	.5	.3	.3	.7	.7	.5

Note: A difference between the two conditions in learning of the controlled AB cluster may evince that differences observed in the original proposed experiment (detailed in Table 1) may be explicable in terms of low-level concurrent extraction of transitional statistics rather than extraction of a single distributional statistic representing average contingency.

Figures

Figure 1. In each block, a series of colored squares will be displayed one after another. Units of the sequence (A, B, C, D…) were represented as the colors of these squares. Possible colors were red, green, blue, cyan, yellow, magenta, brown and black and were randomly assigned to units at the start of each experiment; this assignment was constant throughout blocks. Each colored square was shown in the screen's center for 1000ms; in between each square presentation, a fixation point was shown at the screen's center for 500ms.

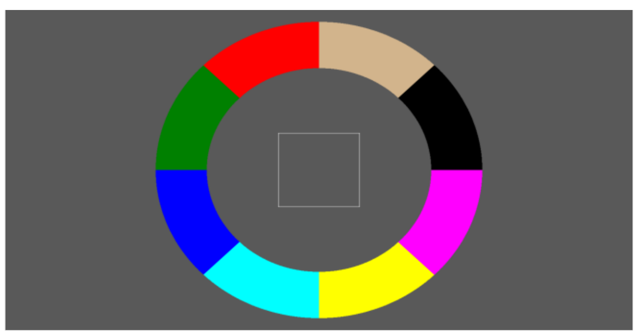

Figure 2. The interface by which participants' extraction of transitional statistics will be tested. In each block, a series of colored squares will be displayed one after another. Randomly (but with enforced spacing) a set number of times within each block, instead of a square, only its outline and a color wheel will be shown. Participants are instructed to move the mouse to the color wheel entry that matches their best guess of the square's color based on their "experience so far".