



Review

Cite this article: Hasson U. 2017

The neurobiology of uncertainty: implications for statistical learning. *Phil. Trans. R. Soc. B* **372**: 20160048.

<http://dx.doi.org/10.1098/rstb.2016.0048>

Accepted: 1 August 2016

One contribution of 13 to a theme issue 'New frontiers for statistical learning in the cognitive sciences'.

Subject Areas:

cognition, neuroscience, behaviour

Keywords:

uncertainty, statistical-learning, entropy, regularity, grammar, language

Author for correspondence:

Uri Hasson

e-mail: uri.hasson@unitn.it; uri@hasson.org

The neurobiology of uncertainty: implications for statistical learning

Uri Hasson

Center for Mind/Brain Sciences, The University of Trento, via delle Regole 101, Mattarello, TN 38123, Italy

UH, 0000-0002-8530-5051

The capacity for assessing the degree of uncertainty in the environment relies on estimating statistics of temporally unfolding inputs. This, in turn, allows calibration of predictive and bottom-up processing, and signalling changes in temporally unfolding environmental features. In the last decade, several studies have examined how the brain codes for and responds to input uncertainty. Initial neurobiological experiments implicated frontoparietal and hippocampal systems, based largely on paradigms that manipulated distributional features of visual stimuli. However, later work in the auditory domain pointed to different systems, whose activation profiles have interesting implications for computational and neurobiological models of statistical learning (SL). This review begins by briefly recapping the historical development of ideas pertaining to the sensitivity to uncertainty in temporally unfolding inputs. It then discusses several issues at the interface of studies of uncertainty and SL. Following, it presents several current treatments of the neurobiology of uncertainty and reviews recent findings that point to principles that serve as important constraints on future neurobiological theories of uncertainty, and relatedly, SL. This review suggests it may be useful to establish closer links between neurobiological research on uncertainty and SL, considering particularly mechanisms sensitive to local and global structure in inputs, the degree of input uncertainty, the complexity of the system generating the input, learning mechanisms that operate on different temporal scales and the use of learnt information for online prediction.

This article is part of the themed issue 'New frontiers for statistical learning in the cognitive sciences'.

1. The role of uncertainty in psychology: a story of waxing and waning (and waxing)

The status of uncertainty (also, entropy, disorder) as an explanatory construct in cognitive psychology has seen fluctuations since the 1950s. Catalysed by Claude Shannon's work and the subsequent advent of information theory, numerous studies from the late 1940s to the early 1960s relied on insights and formalisms of information theory to explain learning, memory and language (for review of early studies, see [1]). To name a few, such studies attempted to quantify the channel capacity of sensory systems by studying the relation between the information in the inputs and the information in behavioural responses, evaluate responses as a function of distributional uncertainty (entropy) or empirically examine stimulus surprise. Some of those studies' insights remain fundamental within computational neuroscience and certain subfields of psychology. However, the limitation of input uncertainty as an explanatory factor was soon realized. In particular, George Miller had already noted in the 1950s [2] that working memory is not constrained by stimulus uncertainty, but by a potential for chunking that depends on long-term familiarity. This, in turn, introduced a strong limitation on the explanatory power of uncertainty as a stand-alone construct in approaches to learning and memory (setting aside more limited applications in domains such as sensory discrimination). Luce [3] covers some of this historical progression and other limitations in his review, 'Whatever happened to information theory in psychology', and Laming [4] offers a detailed

critique of the linking hypotheses that underlie interpretations of behaviour in relation to stimulus uncertainty.

However, the more recent use of non-invasive neurobiological methods, particularly functional magnetic resonance imaging (fMRI), allows a different type of understanding of how humans (and their brains) respond to uncertainty. Whereas the original behavioural studies necessarily relied on quantifying the relation between the information in a stimulus set and the information inherent in responses to that stimulus (e.g. to infer channel capacity), neurobiological studies do not need to rely on overt behaviour. This allows us, for example, to identify brain systems that track input uncertainty and study their capacity, while participants are engaged in behaviours that are unrelated to the stimuli or the manipulation of interest. Studying the neurobiology of uncertainty has become central in several domains, including sensory perception, rule learning (for review, see [5]), the adjustment of predictions [6,7] and understanding anxiety from the perspective of uncertainty [8–10].

Importantly, neurobiological questions about the principles that organize the brain's responses to uncertainty do not necessarily align with, nor are they intended to test, functional accounts of how uncertainty impacts behaviour. In this sense, neurobiological and functional approaches, while at times sharing common concerns, are also concerned with partially separate issues.

2. Neurobiological approaches to uncertainty and statistical learning: interfaces and disconnects

Statistical learning (SL) is a dynamic field that is itself changing in theoretical scope and emphasis. Nonetheless, as a working definition the current discussion adopts Aslin & Newport's [11] perspective where SL is defined as a mechanism that 'enables adults and infants to extract patterns embedded in both language and visual domains'. A similar view is seen in Schapiro and Turk-Browne's neurobiologically focused discussion where SL 'refers to the ability to extract regularities from the environment over time' [12, p. 501].

This intersection between the interest in regularities, on the one hand, and temporally extended learning, on the other hand, delimits the SL domain. It also separates it from related questions such as the coding of instantaneous statistics [13] or distributional learning of continuous variables (e.g. learning the mean and variance of a distribution of certain sensory features [14]). This emphasis naturally leads to the study of nominal rather than continuous variables, which is also consistent with the historical context in which the study of SL was developed. Specifically, SL has been suggested to be a domain-general capacity that underlies language acquisition [15], with an emphasis on the ability to code for specific features of an input stream such as marginal frequencies, transition probabilities and mutual information.

The shared interest in nominal variables and their distribution constitutes an important formal, if not theoretical link between neurobiological approaches to SL and uncertainty (we henceforth focus on neurobiological approaches). At the end of the day, both approaches are interested in learning, but they put different emphasis on what might be learnt. Whereas the theories of SL are focused on how particular regularities or associations between elements are encoded, theories of uncertainty focus on how the overall degree of input uncertainty may be encoded or used for various purposes. Input

uncertainty is a multifaceted construct and can be captured by numerous information-theoretic quantities that differentiate random from non-random inputs. For instance, Shannon's measure of entropy captures (roughly speaking) the relative diversity of input tokens, and Markov entropy reflects the strength of their transition constraints. Other measures, many derived within the field of dynamic systems, load in some way on serial autocorrelation (long-term memory) within the data (e.g. Hurst exponent, attractor dimensions, etc.). Behaviourally, people have been shown to be highly sensitive to such long-term features that impact uncertainty [16,17].

At times, manipulations of uncertainty or statistical structure (as implemented in SL paradigms) amount to a terminological difference. Simple artificial grammars (regular grammars) of the form often used in SL and artificial grammar learning (AGL) paradigms can be represented as a first-order Markov process: one where the next step is conditional only on the current state. When represented this way, the fact that certain transitions are allowed, whereas others are not, produces a quantifiable reduction in uncertainty and distinguishes grammatical from random strings. For this reason, some manipulations used in neurobiological studies of SL are also manipulations of uncertainty. For instance, in an fMRI study, McNealy *et al.* [18] examined the blood-oxygen-level-dependent response when participants heard either random syllable sequences, or syllable sequences generated by a grammar that effectively produced fixed 'words' that could be freely combined. The latter (more regular) condition produced greater activity in lateral temporal cortex (see [19] for a similar paradigm and findings). Thus, most generally, manipulations of uncertainty can identify brain systems sensitive to the statistical structure of the input, and such manipulations may subsume ones pitting random strings against those generated by a grammar. The following sections discuss several theoretical and experimental trends in studying the neurobiology of uncertainty that have implications for studies of SL.

(a) Distributions and associations: macro- and microscale aspects of learning

Manipulations of uncertainty impact *macroscale* properties of a stimulus series (e.g. mean series uncertainty as quantified via Shannon or Markov entropy or any other summary feature related to long-term autocorrelation or embedding dimension of the input). It has been suggested that there may exist brain systems that code for such 'summary statistics'—ones that reduce uncertainty into a single value (see [5] for discussion in context of decision-making under uncertainty). However, uncertainty is also reflected in *microscale* features at the scale of single stimulus tokens, such as how surprising is the appearance of a single token or how much information a certain token provides about what is likely to happen next (cue diagnosticity). In some formalisms, for example, the Rescorla–Wagner model, updating occurs on the microscale—that of transition constraints between pairs of tokens. However, computations operating on both scales likely impact how people code for and respond to uncertainty. Consider a binary [1,0] series derived from the transition structures of the following two processes

- Process 1: $P(1 | 1) = 70\%$; $P(0 | 1) = 30\%$; $P(1 | 0) = 50\%$; $P(0 | 0) = 50\%$
- Process 2: $P(1 | 1) = P(1 | 0) = P(0 | 1) = P(0 | 0) = 50\%$

Series generated from the first process are more regular, because transition constraints are stronger. On some theoretical approaches, brain systems sensitive to this macroscale feature (a ‘summary statistic’) would differentiate the two series for this reason. Several neurobiological studies have modelled brain responses as function of such summary statistics [20–24].

Conjointly, on the microscale of single items, in the regular process, the state marked ‘1’ has higher cue diagnosticity—it provides more information about the next state—than any token in the random process. For this reason, brain systems that establish associations between pairs of tokens may be more frequently engaged for series generated by the first process. Relatedly, systems signalling prediction error may also be more extensively engaged in this case, specifically, when 0 follows 1, which may signal a violated prediction. The hippocampus has been linked to associative learning in continuous series, showing greater activity for tokens highly predictive of a subsequent token [12,25,26]. However, frontoparietal systems have been implicated by studies that explicitly manipulated cue diagnosticity, with greater activity reported for more diagnostic cues [27,28].

The macro vs micro distinction has a parallel in the SL literature concerned with whether adults and children learn macroscale distributional information or, alternatively, knowledge of the bigrams or trigrams whose presence typically differentiates a grammar from random transition network. To illustrate, the fact that infants look longer at elements that make up random series could indicate they differentiate random from regular series owing to sensitivity to a ‘global’ distributional feature, or alternatively, because the regular series contain more frequent cases of particular sequences/transitions. Interestingly, recent work suggests that the latter is the case [29], and the use of neurophysiological measurements may shed further light on this issue in the future. Thiessen *et al.* [30] make a similar distinction between conditional transition probabilities and distributional statistics. Another approach that differentiates micro- and macro-description levels within the context of SL is described in Karuza *et al.* [31]. There, relations between items are considered as edges within a network, and summary features of the network are taken to correlate with the ability to learn and remember the system of interest (in their terminology, these levels correspond to local statistics and complex-network features).

These statistics play an important role in contemporary neurobiological models of language. Understanding whether the brain tracks the information that a linguistic cue provides about potential upcoming ones is becoming increasingly important because of the hypothesis that certain brain systems are engaged in prediction (at the phonemic, morphemic and word level) during comprehension. In particular, it has been suggested that during language comprehension certain brain systems signal the uncertainty of what is likely to appear. This means their activity tracks the variance of the set of potential completions that can follow at each point (formally, Shannon’s entropy of set of immediate complements). There is some empirical support for this proposition, emerging from neuroimaging [32], electrocorticography [33] and magnetoencephalography (MEG) studies [34,35], but these have typically implicated lateral temporal and dorsolateral prefrontal cortex. Technically, it is often possible to dissociate the impact of macro- and macro state features on brain activity by including in explanatory models both the overall macro

state, as well as the microscale features of the stimuli [23]. Nonetheless, the theoretical distinction between micro- and macroscale features is one that requires more empirical investigation. It is unknown whether, in terms of computation, there is a difference between establishing the uncertainty associated with local (short-term) and global (long-term) patterns and whether they rely on different latent capacities. Furthermore, it is unknown whether the coding of local patterns is itself impacted by macro-state features.

This is not to say that studies of uncertainty necessarily carry direct implications for neurobiological models of SL. For instance, some studies have evaluated sensitivity to temporally unfolding features of tonal series such as their fractal properties [36], relative smoothness [37] or magnitude of pitch transitions [38]. Such paradigms do indeed impact the predictability of future stimuli over multiple scales. However, they rely on manipulations of a physical feature (e.g. pitch-changes over time). Consequently, their impact on brain activity may be due not to the coding of statistics of informational features, but also to the way these manipulations impact lower-level sensory processing. For instance, series that consist of smoother or more gradual pitch changes [36,37] could produce stronger neural repetition suppression, because auditory neurons tuned to a certain pitch bandwidth are engaged for longer durations. To conclude, neurobiological studies of uncertainty and SL develop from largely separate theoretical backgrounds. Still, there are interfaces between the two domains, as computational/neurobiological models of uncertainty consider both macro- and microscale features that are of interest to neurobiological models of SL.

(b) From learning to using

Traditionally, behavioural studies of SL have drawn inferences about the outcome of learning from participants’ responses to test trials presented after learning. This has led to sophisticated models of the process [39]. However, more recently, the emphasis on the process of learning has been brought to the fore, as shown for instance in Frost *et al.*’s [15] emphasis on the dynamic aspects of learning wherein internal representations are updated via interactions between current inputs and prior knowledge. In tandem, behavioural studies of SL have begun examining online responses during presentation of the stimulus series [40,41]. Such examinations produce a more direct description of the learning process itself rather than inferring the learning process from responses to test items.

A similar trajectory had occurred in neurobiological studies of AGL, SL and the coding of uncertainty, which have seen a shift from an initial interest in identifying a putative ‘learning system’ to more dynamic ‘real-time’ models of how statistics are acquired and used. The initial emphasis on identifying statistical-learning systems is seen, for instance, in a study that used an AGL paradigm [42] to identify brain regions that differentiated grammatical from ungrammatical test items. The authors concluded that the right caudate tracked adherence to the grammatical rule, whereas hippocampal activity tracked ‘chunk strength’—the extent to which test items were similar to training items. A similar interest in identifying distributional learning systems is seen in early neurobiological studies of uncertainty. Strange *et al.* [23] focused on sensitivity to marginal frequencies and implicated the hippocampus, and two other studies examining sensitivity to transitional probability constraints had linked the hippocampus [21] and left posterior

lateral temporal cortex [20] to these computations. Notably, however, some of these earlier studies were already concerned with how distributional knowledge impacted responses to single tokens—Strange *et al.* [23] quantified the *surprisal* of each item ($-\log(P(x))$) in relation to the distribution of items to that point and found strong stimulus-surprise effects in frontoparietal regions, the thalamus and fusiform cortex (stimuli were visual). Other work [43] had partitioned a putatively initial stage of acquiring statistical knowledge ('process of learning') from later stages where that knowledge is used ('result of learning') and showed they involve different brain systems.

A theoretical shift from learning statistics to 'using statistics' necessarily foregrounds two interrelated processes that are of major interest in current cognitive neuroscience but are seldom addressed within neurobiological studies of SL—the construction of predictions and their evaluations (for exceptions, see [25,26,44,45]). A very large body of neurobiological work is concerned exclusively with these issues, typically outside the context of SL, using paradigms where cue validity is manipulated [28] or where cues are used to indicate future stimuli in different modalities [46]. This work has been reviewed extensively [7] and we will not do so here. What is important is that during the process of SL, brain activity inevitably reflects a combination of computations related to updating of distributional knowledge, as well as the construction of predictions licensed by that knowledge and their evaluation. Construing SL from an integrative perspective that takes into account the implicit detection of patterns or distributional information, and the subsequent use of this knowledge for prediction offers multiple opportunities for theoretical advances and may likely lead to different conclusions than those drawn from work to date. To illustrate, in cue–target paradigms, cues with higher validity (i.e. ones more informative about a future target) evoke greater activity than less valid cues in frontoparietal regions [28]. However, when regular and random inputs are presented for *passive* observation in tasks that do not demand executive function, then the converse pattern is found: stimuli series where (on average) transition constraints are higher are associated with *less* activity in frontoparietal regions than random series [47]. It has also been shown that when a stimulus stream conveys predictable knowledge in separate dimensions (e.g. shape and colour), then regularities are associated with less activity (than a random condition) for the non-attended dimension, but more activity than a random condition for the attended dimension [43]. In other work [48], it was shown that brain responses to predictable stimuli are themselves impacted by attention allocated to the input stream: when predictable stimuli were presented at an attended screen side, they were associated with increased activity when compared with a condition where no prediction was possible. However, the opposite pattern was found for stimuli presented at an unattended screen side—there, predictable stimuli evoked less activity than ones for which predictions were not viable. Thus, the online coding of regularities, as computed in the context of implicit paradigms that are of interest to SL, may shed new light on how learning and prediction occur in natural contexts.

(c) The importance of temporal scope

At present, one difference between neurobiological studies of SL and uncertainty pertains to their theoretical emphasis on the temporal constants over which information is integrated

during learning. SL developed as a potential explanation for language acquisition [15], and the two capacities are often seen as intertwined and potentially loading on the same latent capacities [49]. An important feature of language statistics (e.g. phonotactic frequencies or transition probabilities between phonemes or syllables) is that these form a relatively stationary system. In other words, it is unlikely that once phonotactic or syllabic-level transition constraints are acquired by an adult, those would need to be strongly modified as a result of subsequent exchanges in that language. In this respect, a language system is largely a 'fixed target' where an assumption of stationarity is licensed. (There may be few exceptions that prove the rule, for instance, circumstances of second language learning in contexts where L1 and L2 share a syllabic/phonetic inventory). This implicit assumption of stationarity is consistent with the finding that individuals show gradually increasing (or decreasing) responses to stimuli drawn from regular (but not random) series, which has been documented in several studies [18,19], and with fMRI work [43] that separated an initial phase reflecting the 'process of learning' from a later phase of 'result of learning'.

The stationarity assumption was also implicit in earlier models of the coding of uncertainty, which assumed that individuals code statistics as an ideal Bayesian observer—one that weighs new information in relation to the history of *all* previous trials. In practice, this meant analysing entire stimulus blocks as a single condition [20], or modelling brain activity on a current trial, Trial(*n*), in relation to the distribution of all input trials experienced to that point (from Trial1 to Trial(*n*-1)) [21,23]. However, the weakness of this assumption was soon evident. First, even when events are drawn from a stationary distribution, people appear to attribute more weight to the most recent trials. This was elegantly demonstrated by Huettel *et al.* [50]. The authors reported fMRI data indicating that individuals tend to 'perceive patterns in random series'. Specifically, when observing a binary pattern generated by a random process, individuals were sensitive to the interruption of 'local streaks' that were as short as two to three repetitions or five to six alternations. Thus, even though the system was stationary and random so that participants could not make any prediction of the future, they were still sensitive to statistical structure in the very recent past. Similarly examining stationary cases, work within a classical conditioning paradigm [51] has shown that different brain areas learn over different temporal constants as reflected in their activity profiles having different learning rates when modelled via a Rescorla–Wagner model. In another neuroimaging study, Harrison *et al.* [52] presented participants with visual stimuli generated by a stationary first-order Markov process. They modelled responses to each stimulus assuming there exists a limited memory capacity that constrains the ability to establish associative relations between token pairs (this capacity was estimated from a behavioural study). They found that frontal medial cortices (anterior cingulate, ACC) have longer integration windows than primary visual cortices, with the latter losing almost all traces of events beyond the last four trials. Bornstein & Daw [53] addressed a similar question, presenting participants with a stationary Markov process consisting of four unique images whose transition-probabilities were controlled. Their modelling of behavioural responses to these stimuli indicated two superimposed learning processes associated with two different learning rates (faster rate = 0.5 and slower rate = 0.1). Given

these estimates, they constructed model-informed predictors of brain activity during series presentation that reflected, at each point in the series, the uncertainty about the next stimulus. They constructed such model-informed regressors assuming both slow and fast learning rates. The model with the fast learning rate accounted for activity in the ventral striatum and insula, whereas the model with a slower learning rate accounted for activity in the anterior hippocampus. To summarize, all these studies show that even in a stationary context where features of the sampling distribution do not change over time, people do not function as ideal observers, and as Bornstein and Daw wrote (p. 1014), ‘this pattern excludes models that do not incorporate forgetting of past experience’. The reliance on prior input can be directly linked to working memory capacity (WMC) [44]: individuals with higher WMC benefit more strongly from the presence of regularities when performing a behavioural task. Furthermore, when comparing MEG (theta band) activity prior to appearance of stimuli in random and regular series, higher WMC is linked to stronger differentiation between these two conditions indicating greater sensitivity to the statistical context.

Complementary work on temporally bounded integration has identified brain systems implicated in coding statistics of non-stationary streams. Initial work [54] conducted within a reward-based reinforcement learning paradigm (a task that strongly relies on an executive component) showed that individuals calibrate their learning rate—the degree to which they update prior information based on the incoming one—to match the volatility of the environment. More volatile contexts produce faster learning rates, and neurobiologically, activity in the ACC tracked the volatility of the environment that itself changed over time. In another fMRI study that focused on passive perceptual learning [55], participants listened to tonal series where the transition constraints between four tones changed continuously over timescales of 10 s, following a preset profile of gradual increases and decreases. The study documented learning on two scales: some regions (mainly perisylvian) tracked the level of regularity in the recent 10 s, whereas other regions tracked a slower-scale process characterized by gradual increases or decreases in regularity over time.

All this suggests some incongruity between the computational demands involved in tracking the statistical nature of events in a continuously changing world and those involved in acquiring the statistics of more stationary linguistic systems. Formally, this distinction is similar to one made in the reinforcement learning literature between two types of uncertainty [56]: first, the *known uncertainty* of a set of potential outcomes given the current stimulus (e.g. the uncertainty captured by Shannon’s entropy of the set of known possible outcomes), assuming veridical knowledge of the contingency structure in a system. Second, the *unknown uncertainty*, which reflects lack of veridical knowledge about what the contingency structure in the environment actually is; this knowledge is reduced in volatile environments but can approach certainty in stationary ones. (These have also been referred to as outcome uncertainty and rule uncertainty [5].) Language learners may be justified in assuming that language statistics are associated with low volatility, so that rule uncertainty is low, whereas understanding the environment may entail a different calibration of such parameters since rule-uncertainty is higher.

It may be that sensitivity to statistical structure on short timescales underlies a more general process of segmenting

continuous inputs into different phases or events. As shown initially by Zacks *et al.* [57], there exists a brain network consisting of mainly occipital–parietal regions that shows transient activation increases at boundaries between naturally unfolding events. This network has been implicated in event segmentation during movie viewing and written text comprehension (see [58] for review). Yet the same network has been implicated in tracking changes in statistical structure in tonal series lacking any semantics [59].

To summarize, studies examining temporal integration windows suggest dissociable operations that occur, in parallel, over longer and shorter timescales. One possibility is that integration over short temporal constants is useful for monitoring changes in non-stationary environments, whereas integration occurring over longer epochs is useful for making more precise predictions in stationary environments. The similarity of networks linked to subjective changes in statistical structure and those found for event segmentation suggests that integration over short temporal constants is related to more general capacities of input segmentation.

(d) The special status of patterns

As seen in the definitions of SL cited above, one of its main aims is understanding the computational and neurobiological mechanisms that allow for learning regularities or patterns. From this perspective, *structure*—instantiated either through statistical (stochastic) constraints or through (deterministic) sequences—stands in contrast to the ‘default’ or non-marked random case. In neurobiological studies of SL, this is expressed in designs that contrast regular and random series [18,19]. The contrast between item-pairs associated with high versus low mutual information [26] can also be seen as reflecting similar interests at the micro level. Most generally, this conceptual framework and the experimental paradigms it derives assume a monotonic relation between associative structure and brain activity. This assumption also underlies many neurobiological studies of uncertainty that set out from the premise that brain regions sensitive to uncertainty will track it monotonically. In practice, this means setting up binary contrasts between regular and random conditions in block designs [18,20] or, in studies conducted in the Bayesian framework, modelling responses to each stimulus as a *linear* function of input entropy.

However, an emerging picture from both behavioural and neurobiological studies is that the *a priori* assumption of a linear or monotonic relation between input structuredness and brain activity results in a partial understanding of computations related to statistical processing and their neurobiological basis. Specifically, as reviewed below, some brain regions are involved in statistical computations but do not differentiate highly regular from random series.

Cognitive psychologists have already suggested that people may track uncertainty, though non-monotonically. An example is Loewenstein’s ‘information-gap’ model of curiosity [60]. In this model, curiosity is a function of (i) a person’s assessment of the maximal uncertainty a system may exhibit [MaxU], (ii) their current level of uncertainty [CurrentU] and (iii) their target criterion for uncertainty [MinU; 0 if one aims for veridical knowledge]. On this formalization, curiosity, or the impetus for exploration, is defined as the information gap, $[(\text{CurrentU} - \text{MinU}) / (\text{CurrentU} - \text{MaxU})]$. Assuming that people strive to minimize uncertainty (fixing MinU = 0), this means that curiosity and its resulting actions will not

directly track environmental uncertainty, as MaxU is equally important. Put differently, even given the same level of input uncertainty, people will show greater curiosity for inputs generated by systems whose maximal uncertainty is higher. On this account, there is no direct relation between input uncertainty and curiosity, though all else being equal, increased uncertainty should induce greater curiosity.

The connections between randomness, complexity and cognition have long been of interest in psychology (for recent treatment, see [61]). There are several psychological models that produce U-shaped responses as uncertainty increases; we have reviewed these in prior work [62] and will not do so here (see [63] for early, exemplary behavioural study and [1] for review of early work). Theoretical work within complexity science also suggests that some uncertainty-related computations produce the highest values for series that are neither highly ordered nor random, but are instead associated with mid-levels of uncertainty. This points to an important distinction between randomness and complexity. The basic premise of these approaches is that complexity is not a monotonic function of randomness and that these two concepts should be differentiated. The emphasis of some complexity theories on the systems that generate the input (rather than the input itself) is also sensible from a psychological perspective, as it speaks to a different level of generalization than is typically studied. Consider for instance a regular input that consists of four tokens with very strong transition constraints, which thereby differentiate it from a matched random process. There may exist cognitive systems that construct and maintain a representation akin to this 4×4 transition matrix, and update this model throughout learning (this assumption underlies several learning models, such as the Rescorla–Wagner model or Bayesian approaches that assume a representation via multivariate distributions). However, this level of description assumes that input frequencies are retained faithfully. Nonetheless, a more abstract, sparse description of the generating system can be formulated: namely in the regular series *some tokens only follow other tokens*, whereas in the irregular series *each token follows all others equally often*. Representations at this abstraction level suffice for knowing when a regular state shifts to a random one and vice versa, but do not rely on veridically retaining quantitative distributional information such as which transitions are strong or weak, how many different tokens exist, or even summary statistics such as the exact uncertainty level of the most ordered series. As mentioned, such descriptions are lengthier and more difficult to formulate for inputs with mid-levels of regularity. Other formal approaches, conceptually similar to Leowenstein’s information-gap model, treat complexity as a combined (weighted) function of a system’s distance from its completely ordered and completely disordered possible states, with certain weight-combinations producing an inverse-U output as function of uncertainty (see [64] for accessible treatment).

Several neuroimaging studies have identified brain systems that track input uncertainty either quadratically (i.e. with a curvilinear, U-shaped profile) or nonlinearly. In one study, participants heard long tonal series that varied parametrically in regularity [62]. This study showed that whole-brain connectivity between the ACC and several brain regions followed regularity via a quadratic trend (a similar finding was shown for the hippocampus). Another study [55] found that activity in lateral temporal regions tracked the strength of transition constraints among auditory tones with a quadratic response profile. In another study [47], participants observed visual

series where the location or semantic category of the next image was or was not predictable. This produced three levels of uncertainty: a baseline condition where neither dimension was predictable (maximum uncertainty), two ‘single-regularity’ conditions where either location or category were predictable (medium uncertainty) and a ‘dual regularity’ condition where both dimensions were predictable (low uncertainty). Interestingly, the study identified several brain regions that tracked uncertainty quadratically; the conditions with medium-uncertainty produced less activity than either the low-uncertainty and maximal-uncertainty conditions.

While the discussion above focused on macroscale features, quadratic responses may additionally reflect sensitivity to microscale features of a system. As already noted by Hebb [65, p. 149], ‘up to a certain point, lack of correspondence between expectation and perception may simply have a stimulating (or ‘pleasurable’) effect’. Developing this intuition within the SL approach, Kidd *et al.* [41] studied surprisal-correlated behavioural indices on the single trial level within continuous series. Analysing children’s looking times throughout the series, they found that children were less likely to look away from images whose surprisal was intermediate—neither too low or too high. To the extent that series with moderate levels of regularity (on average) also contain a greater proportion of trials with moderate surprisal, this could also explain why such series are associated with greater activity.

Why should neurobiological theories of SL or uncertainty be concerned about non-monotonic responses to structuredness, regularity or uncertainty? From a strictly neurobiological perspective, if the goal is to identify brain systems that are sensitive to statistical structure—in the sense that they differentiate random from non-random series—then it is already clear that some brain areas may satisfy this criteria *even though they do not differentiate highly regular from random series*. This, practically, calls for using parametric modulations rather than binary contrasts between regular and random series. A similar argument for parametric manipulations has been made in studies of decision-making under uncertainty [5] where it has been argued that the comparison between certain and (almost) certain states may load on brain states not directly related to the coding of uncertainty. However, the computational/functional concerns are equally important, as existing data show that computational models that explain how patterns or structures are identified and memorized (the SL perspective) cannot easily account for such trends. Monotonic responses are highly consistent with the perspective of compression and memory, because more random inputs are, by definition, less compressible, offer less opportunities for chunking [66], and therefore impose greater memory demands. They are also consistent with theories in which anxiety scales linearly with uncertainty [10]. In contrast, quadratic brain responses with increasing uncertainty reflect different computations, and an important goal of future work is to expose their characteristics, which at the current point remain largely speculative.

3. Current explanatory principles meet challenges from data, and desiderata for future developments

Section 2 reviewed issues that are of shared concern to computational/neurobiological approaches to uncertainty and SL.

This section focuses more specifically on current neurobiological approaches to uncertainty, how they ‘carve up’ the computations that need to be explained, and on their relation to neurobiological models of SL. It then highlights several challenges to such accounts from recent findings. Based on these, it outlines several requirements for future neurobiological frameworks focused on uncertainty.

(a) Organizing principles of current approaches to neurobiology of uncertainty

There exists a substantial body of neurobiological research on uncertainty, developed mainly within reinforcement learning studies and research into decision-making under uncertainty (particularly economic decisions). In a review of that literature, Bach & Dolan [5] proposed a useful distinction between several different types of uncertainty for guiding neurobiological work: (i) sensory uncertainty that reflects uncertainty about physical features of the stimulus; (ii) uncertainty about present state (e.g. are we in street x or street y); (iii) outcome uncertainty: given the current state, what are the probabilities of the set of potential outcomes; and (iv) rule uncertainty: what is the uncertainty about state-outcomes contingencies in and of themselves. Bach and Dolan suggest different regions are sensitive to different types of uncertainty. With relation to SL, one can say that typical statistical-learning challenges as examined in adult populations are associated with low sensory uncertainty as sensory stimuli are typically constructed to be unambiguous and easily discriminable among themselves. Furthermore, rule uncertainty is low in that contingencies are rarely changed throughout a study, and participants may be instructed in a way that conveys that these contingencies do not change (see [67] for the difficulty of acquiring a second rule system). Outcome uncertainty, however, loads on the quantities of marginal frequencies, transition probabilities and mutual information that are core to SL. The applicability of research on outcome uncertainty within decision-making paradigms may be limited though; such neurobiological investigations are often strongly related to the monetary gain or loss associated with each possibility (where expected utility is defined as [value \times probability]). This is an element tangential to the phenomena SL studies are typically interested in.

Ma & Jazayeri [68] present a computational/neurobiological approach to uncertainty that differs from that of Bach and Dolan’s in that it aims to explain one’s own actions and their potential utility. This framework also partitions between sensory and other types of uncertainty. Beyond sensory uncertainty (e.g. ‘is this liquid milk’) a different level of uncertainty pertains to the applicability of different sensorimotor contingencies (e.g. ‘if it is milk, then should I drink it’). This, in turn, is linked to a set of potential motor actions (with varying probabilities), and finally, an action-linked reward that is also assigned a probability. The authors discuss neurobiological systems that code for these sorts of uncertainties—each is thought to be linked to a belief distribution, and the framework explains how neuronal populations can represent such prior beliefs in a Bayesian framework. This work constitutes an important position on computational principles that may govern processing of incoming information given prior knowledge.

Other frameworks that inform work on uncertainty deal with the coding of sequential structures. Friston & Buzsaki [69] suggest that hippocampus maintains information about sequential transitions (e.g. where/when or what/when), and

interacts with cortical systems to represent the content of the next event. Dehaene *et al.* [70] approach the related issue of coding for serial order. They partition between different types of sequential information and their putative neurobiological underpinnings. Particularly relevant are their explanations for coding transition structure and chunking. For transition structure, their model emphasizes the importance of different *sensory cortices*. They emphasize that sensory regions coding for physical features of a stimulus are implicated in anticipatory predictions and production of prediction-error terms. For instance, sensory cortices code for the surprisal of sensory stimuli—for tonal stimuli, the last ‘B’ in ABABB evokes a surprise response in auditory cortex rather than one showing repetition-suppression, reflecting a violation of expectation. Chunking—the joining of elements into a fixed configuration—is taken to be mediated by a different mechanism than the one mediating lower-level mismatch responses, with the prototypical case being ‘word chunks’ in typical SL paradigms (which are characterized by very high pointwise mutual information). The authors link these to left auditory association cortex or left inferior frontal gyrus.

(b) The importance of non-general processes

While computational models of uncertainty outline generic computations, they make no clear commitments about how these may be instantiated in neurobiological substrates. What is already evident is that ‘unified approaches’ that posit single biological systems that code for sequential structure or prediction error are no longer viable. Setting aside the impact of sensory uncertainty, uncertainty about actions or reward, even the core issue of outcome uncertainty—what are the potential future events given the present—is one whose neurobiology is poorly understood. As mentioned, prediction error appears to be generated in sensory-specific systems, and even within a single modality, the statistical structure of different sorts of dimensions (e.g. location, category) may be coded in partially different systems [47]. It is therefore unclear whether a single brain system codes for the regularity of input streams in different sensory modalities or is generally involved in generating predictions for those. It may be, however, that different sensory systems implement similar computations over inputs with different features (see [71] for potential computational implementation). Indeed, a recent approach to SL by Frost *et al.* suggests that it is a ‘set of domain-general computational principles that operate in different modalities’ and therefore reflects modality-specific constraints [15]. As outlined below, sensitivity to uncertainty appears to be strongly determined by constraints that are not only related to sensory features, but are fundamentally determined by the dimension being tracked, even within a sensory modality. Furthermore, a more difficult challenge arises when considering the brain as a functional network whose core features may change with the level of uncertainty. To make these issues concrete, the four points below exemplify the sort of data that future neurobiological theories of uncertainty would need to account for.

(i) Lack of support for modality-independent or modality-encapsulated processing

An fMRI study [22] in which participants were presented with auditory or visual series that varied across four levels of uncertainty found that different neural systems were involved in

tracking uncertainty in visual and auditory series. No region was sensitive to uncertainty in both modalities. Another fMRI study [72] examined sensitivity to uncertainty in auditory, visual and audiovisual inputs, and found that the audiovisual condition strongly altered sensitivity to regularity in bilateral primary auditory cortices; these regions tracked regularity in the auditory, but not audiovisual condition. Thus, conclusions drawn about the neurobiological basis of uncertainty that are based on studying inputs in one modality will not necessarily generalize to alternative modalities or even multimodal contexts.

(ii) Fundamental role for input familiarity

Sensitivity to input-uncertainty is impacted by familiarity with the input tokens. In one study [24], auditory series that varied in uncertainty consisted of either familiar tokens or unfamiliar tokens (syllables in the participants' native language or bird chirps, respectively). In the syllable-streams, sensitivity to uncertainty was found within 4 s from series onset, but in this early period, there was no sensitivity to disorder in the bird-chirp streams (these effects were limited to low-level auditory regions). Conversely, in other regions, brain activity after the series' end (13 s from onset) tracked uncertainty for the bird-chirp series, but not the syllable series. Behavioural data suggested this might be related to more accurate segmentation of syllable streams into constituent units. The issue of segmentation may be a greater concern within auditory streams, where elements lack a natural physical boundary (between phonemes or words) when compared with visual streams where stimuli are typically separated by blanks.

(iii) Indications that future-uncertainty is a non-unitary construct

Outcome uncertainty refers to the overall uncertainty about a future event (or events) at a certain point. For instance, given a stream such as 11001110011110 one can ask what is the likelihood of the next event. In this specific example, two forms of information are evident. On the one hand, 1 is more frequent, but on the other hand, the first 0 after a 1 is always followed by a 0. Thus, marginals and transition constraints provides different sorts of information. Whether or not different systems track these sources of information was addressed by an fMRI study [55] where these forms of uncertainty were manipulated orthogonally within a single tone series. Participants listened to a 10 min tonal series that was constructed to satisfy three criteria: (i) the transition constraints between tones, quantified via Markov entropy, fluctuated over time via a predefined specification when quantified over 10 s windows; (ii) the relative diversity of the tokens, quantified via Shannon entropy similarly fluctuated over this timescale; (iii) fluctuations in Markov entropy and Shannon entropy were orthogonal (i.e. uncorrelated) over the entire tonal series. The study found that different brain regions were sensitive to these two facets of uncertainty. The dissociation between systems sensitive to transition constraints between tokens and those sensitive to overall diversity is supported by computational models of SL [73], which draw a similar distinction between two statistical processes that putatively govern SL: extraction of patterns and integration of information to arrive at global features.

(iv) Evidence that whole-brain connectivity structure fundamentally changes with level of uncertainty

Current neurobiological approaches to uncertainty are 'fixed' in the sense that they assume that inputs/environments with

different statistics are parsed by the same brain network (a 'fixed organization' model). This holds also for the computational frameworks reviewed above, which assume that certain regions code for distributions of prior information, independent of those distributions' parameters. A fixed organization view implicitly assumes that certain regions or local networks perform computations related to input statistics such as the representation of distributional features (e.g. summary statistics), associative binding, prediction and updating of knowledge based on prediction error. Crucially, the operating characteristics or connectivity of these regions do not themselves change with input statistics. This perspective leads to a specific sort of description for brain activity, for example, focusing on a specific region's activity over the course of learning of inputs that differ in statistical features [18,19]. In contrast, a 'network reorganization' perspective allows for qualitative changes in whole-brain organization as a function of input statistics. There is some recent work suggesting that statistical context (and context more generally) can impact core organizational features of brain connectivity. An fMRI study that examined network connectivity while listening to tonal series that varied in regularity [74] found evidence for changes in core topological features of these networks. When contrasting network organization during listening to random versus highly regular series, network modularity was lower for highly regular series, and the actual network partition structure also differed significantly between these conditions. Other work has shown that, more generally, task demands very strongly impact the organization of functional networks [75], and documented a relation between the extent of such reorganization and behavioural efficiency. These are likely to be central themes in future work.

4. Summary and future directions

Neurobiological accounts of uncertainty originate from a different theoretical background than those of SL. Yet, they convey useful lessons for the development of neurobiological theories of SL. They suggest that a comprehensive understanding of how the brain codes for statistics would benefit from the following: (i) greater appreciation of modality-specific coding constraints, (ii) reduced emphasis on pattern learning, accompanied by stronger focus on the import of non-monotonic responses to structure, (iii) increased emphasis on the dynamics of the learning process itself including its interaction with mechanisms related to generation and evaluation of predictions, and (iv) a more serious attempt to study the brain as a network whose configuration and mode of operation can change markedly with input statistics. One limitation of the current state of the art, which is likely owing to the relative novelty of the fields of study in question, is the relative fragmentation of theoretical perspectives having to do with SL, uncertainty and online expectation and prediction. To demonstrate, consider four recent reviews of the neural basis of SL [12], prediction and expectation [6], estimation of uncertainty [5] and the role of expectation in perception [7]. As a proxy for shared perspective, we calculated the overlap between cited references for each pair of reviews, quantified via the Jaccard index (shared references/total sum). The mean value of all eight pairwise analyses was 1.7% (s.d. = 2%; range = 0–4%). A holistic approach that treats all these as components of a unified 'statistical learning' or 'uncertainty learning' computation and examines those conjointly may lead

to exciting discoveries that cannot be predicted from approaches that deal with each computation separately.

Competing interests. The author has no competing interests.

Funding. This work was partly supported by a European Research Council starting grant (ERC-STG no. 263318 NeuroInt) to U.H.

Acknowledgements. The author thanks Sam Nastase and Michael Tobia for helpful comments on the ideas presented here and two anonymous referees for very constructive and detailed suggestions. Some of the ideas related to the importance of temporal scope emerged during a related discussion with Dick Aslin.

References

- Garner WR. 1962 *Uncertainty and structure as psychological concepts*. Oxford, UK: Wiley.
- Miller GA. 1956 The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* **63**, 81–97. (doi:10.1037/h0043158)
- Luce RD. 2003 Whatever happened to information theory in psychology? *Rev. Gen. Psychol.* **7**, 183–188. (doi:10.1037/1089-2680.7.2.183)
- Laming D. 2010 Statistical information and uncertainty: a critique of applications in experimental psychology. *Entropy* **12**, 720–771. (doi:10.3390/e12040720)
- Bach DR, Dolan RJ. 2012 Knowing how much you don't know: a neural organization of uncertainty estimates. *Nat. Rev. Neurosci.* **13**, 572–586. (doi:10.1038/nrn3289)
- Schubotz RI. 2015 Prediction and expectation. In *Brain mapping: an encyclopedic reference* (ed. AW Toga), pp. 295–302. San Diego, CA: Academic Press.
- Summerfield C, de Lange FP. 2014 Expectation in perceptual decision making: neural and computational mechanisms. *Nat. Rev. Neurosci.* **15**, 745–756. (doi:10.1038/nrn3838)
- Herry C, Bach DR, Esposito F, Di Salle F, Perrig WJ, Scheffler K, Luthi A, Seifritz E. 2007 Processing of temporal unpredictability in human and animal amygdala. *J. Neurosci.* **27**, 5958–5966. (doi:10.1523/JNEUROSCI.5218-06.2007)
- Bach DR, Seifritz E, Dolan RJ. 2015 Temporally unpredictable sounds exert a context-dependent influence on evaluation of unrelated images. *PLoS ONE* **10**, e0131065. (doi:10.1371/journal.pone.0131065)
- Hirsh JB, Mar RA, Peterson JB. 2012 Psychological entropy: a framework for understanding uncertainty-related anxiety. *Psychol. Rev.* **119**, 304–320. (doi:10.1037/a0026767)
- Aslin RN, Newport EL. 2012 Statistical learning: from acquiring specific items to forming general rules. *Curr. Dir. Psychol. Sci.* **21**, 170–176. (doi:10.1177/0963721412436806)
- Schapiro A, Turk-Browne N. 2015 Statistical learning. In *Brain mapping: an encyclopedic reference* (ed. AW Toga), pp. 501–506. San Diego, CA: Academic Press.
- Chong SC, Treisman A. 2003 Representation of statistical properties. *Vision Res.* **43**, 393–404. (doi:10.1016/S0042-6989(02)00596-5)
- Summerfield C, Behrens TE, Koehlin E. 2011 Perceptual classification in a rapidly changing environment. *Neuron* **71**, 725–736. (doi:10.1016/j.neuron.2011.06.022)
- Frost R, Armstrong BC, Siegelman N, Christiansen MH. 2015 Domain generality versus modality specificity: the paradox of statistical learning. *Trends Cogn. Sci.* **19**, 117–125. (doi:10.1016/j.tics.2014.12.010)
- Smithson M. 1997 Judgment under chaos. *Organ. Behav. Hum. Decis. Process.* **69**, 58–66. (doi:10.1006/obhd.1996.2672)
- Stephen DG, Dixon JA. 2011 Strong anticipation: multifractal cascade dynamics modulate scaling in synchronization behaviors. *Chaos Solitons Fractals* **44**, 160–168. (doi:10.1016/j.chaos.2011.01.005)
- McNealy K, Mazziotta JC, Dapretto M. 2006 Cracking the language code: neural mechanisms underlying speech parsing. *J. Neurosci.* **26**, 7629–7639. (doi:10.1523/JNEUROSCI.5501-05.2006)
- Cunillera T, Camara E, Toro JM, Marco-Pallares J, Sebastian-Galles N, Ortiz H, Pujol J, Rodriguez-Fornells A. 2009 Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage* **48**, 541–553. (doi:10.1016/j.neuroimage.2009.06.069)
- Bischoff-Grethe A, Proper SM, Mao H, Daniels KA, Berns GS. 2000 Conscious and unconscious processing of nonverbal predictability in Wernicke's area. *J. Neurosci.* **20**, 1975–1981.
- Harrison LM, Duggins A, Friston KJ. 2006 Encoding uncertainty in the hippocampus. *Neural Netw.* **19**, 535–546. (doi:10.1016/j.neunet.2005.11.002)
- Nastase S, Iacovella V, Hasson U. 2014 Uncertainty in visual and auditory series is coded by modality-general and modality-specific neural systems. *Hum. Brain Mapp.* **35**, 1111–1128. (doi:10.1002/hbm.22238)
- Strange BA, Duggins A, Penny W, Dolan RJ, Friston KJ. 2005 Information theory, novelty and hippocampal responses: unpredicted or unpredictable? *Neural Netw.* **18**, 225–230. (doi:10.1016/j.neunet.2004.12.004)
- Tremblay P, Baroni M, Hasson U. 2013 Processing of speech and non-speech sounds in the supratemporal plane: auditory input preference does not predict sensitivity to statistical structure. *Neuroimage* **66**, 318–332. (doi:10.1016/j.neuroimage.2012.10.055)
- Reddy L *et al.* 2015 Learning of anticipatory responses in single neurons of the human medial temporal lobe. *Nat. Commun.* **6**, 8556. (doi:10.1038/ncomms9556)
- Turk-Browne NB, Scholl BJ, Johnson MK, Chun MM. 2010 Implicit perceptual anticipation triggered by statistical learning. *J. Neurosci.* **30**, 11 177–11 187. (doi:10.1523/JNEUROSCI.0858-10.2010)
- Cristescu TC, Devlin JT, Nobre AC. 2006 Orienting attention to semantic categories. *Neuroimage* **33**, 1178–1187. (doi:10.1016/j.neuroimage.2006.08.017)
- Egner T, Monti JM, Trittschuh EH, Wieneke CA, Hirsch J, Mesulam MM. 2008 Neural integration of top-down spatial and feature-based information in visual search. *J. Neurosci.* **28**, 6141–6151. (doi:10.1523/JNEUROSCI.1262-08.2008)
- Addyman C, Mareschal D. 2013 Local redundancy governs infants' spontaneous orienting to visual-temporal sequences. *Child Dev.* **84**, 1137–1144. (doi:10.1111/cdev.12060)
- Thiessen ED, Kronstein AT, Hufnagle DG. 2013 The extraction and integration framework: a two-process account of statistical learning. *Psychol. Bull.* **139**, 792–814. (doi:10.1037/a0030801)
- Karuzza EA, Thompson-Schill SL, Bassett DS. 2016 Local patterns to global architectures: influences of network topology on human learning. *Trends Cogn. Sci.* **20**, 629–640. (doi:10.1016/j.tics.2016.06.003)
- Willems RM, Frank SL, Nijhof AD, Hagoort P, Van den Bosch A. 2015 Prediction during natural language comprehension. *Cereb. Cortex.* **26**, 2506–2516. (doi:10.1093/cercor/bhv075)
- Cibelli ES, Leonard MK, Johnson K, Chang EF. 2015 The influence of lexical statistics on temporal lobe cortical dynamics during spoken word listening. *Brain Lang.* **147**, 66–75. (doi:10.1016/j.bandl.2015.05.005)
- Ettinger A, Linzen T, Marantz A. 2014 The role of morphology in phoneme prediction: evidence from MEG. *Brain Lang.* **129**, 14–23. (doi:10.1016/j.bandl.2013.11.004)
- Fruchter J, Linzen T, Westerlund M, Marantz A. 2015 Lexical preactivation in basic linguistic phrases. *J. Cogn. Neurosci.* **27**, 1912–1935. (doi:10.1162/jocn_a_00822)
- Patel AD, Balaban E. 2000 Temporal patterns of human cortical activity reflect tone sequence structure. *Nature* **404**, 80–84. (doi:10.1038/35003577)
- Overath T, Cusack R, Kumar S, von Kriegstein K, Warren JD, Grube M, Carlyon RP, Griffiths TD. 2007 An information theoretic characterisation of auditory encoding. *PLoS Biol.* **5**, e288. (doi:10.1371/journal.pbio.0050288)
- Schubotz RI, von Cramon DY. 2002 Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: an fMRI study. *Neuroimage* **15**, 787–796. (doi:10.1006/nimg.2001.1043)
- Pothos EM. 2010 An entropy model for artificial grammar learning. *Front. Psychol.* **1**, 16. (doi:10.3389/fpsyg.2010.00016)

40. Kidd C, Piantadosi ST, Aslin RN. 2014 The Goldilocks effect in infant auditory attention. *Child Dev.* **85**, 1795–1804. (doi:10.1111/cdev.12263)
41. Kidd C, Piantadosi ST, Aslin RN. 2012 The Goldilocks effect: human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE* **7**, e36399. (doi:10.1371/journal.pone.0036399)
42. Lieberman MD, Chang GY, Chiao J, Bookheimer SY, Knowlton BJ. 2004 An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *J. Cogn. Neurosci.* **16**, 427–438. (doi:10.1162/089892904322926764)
43. Aizenstein HJ, Stenger VA, Cochran J, Clark K, Johnson M, Nebes RD, Carter CS. 2004 Regional brain activation during concurrent implicit and explicit sequence learning. *Cereb. Cortex* **4**, 199–208. (doi:10.1093/cercor/bhg119)
44. Cashdollar N, Ruhnau P, Weisz N, Hasson U. 2016 The role of working memory in the probabilistic inference of future sensory events. *Cereb. Cortex.* (doi:10.1093/cercor/bhw138)
45. Schapiro AC, Kustner LV, Turk-Browne NB. 2012 Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr. Biol.* **22**, 1622–1627. (doi:10.1016/j.cub.2012.06.056)
46. Langner R, Kellermann T, Boers F, Sturm W, Willmes K, Eickhoff SB. 2011 Modality-specific perceptual expectations selectively modulate baseline activity in auditory, somatosensory, and visual cortices. *Cereb. Cortex* **21**, 2850–2862. (doi:10.1093/cercor/bhr083)
47. Davis B, Hasson U. 2016 Predictability of what or where reduces brain activity, but a bottleneck occurs when both are predictable. *Neuroimage.* (doi:10.1016/j.neuroimage.2016.06.001)
48. Kok P, Rahnev D, Jehee JF, Lau HC, de Lange FP. 2012 Attention reverses the effect of prediction in silencing sensory signals. *Cereb. Cortex* **22**, 2197–2206. (doi:10.1093/cercor/bhr310)
49. Misyak JB, Christiansen MH. 2012 Statistical learning and language: an individual differences study. *Lang. Learn.* **62**, 302–331. (doi:10.1111/j.1467-9922.2010.00626.x)
50. Huettel SA, Mack PB, McCarthy G. 2002 Perceiving patterns in random series: dynamic processing of sequence in prefrontal cortex. *Nat. Neurosci.* **5**, 485–490. (doi:10.1038/nn841)
51. Glascher J, Buchel C. 2005 Formal learning theory dissociates brain regions with different temporal integration. *Neuron* **47**, 295–306. (doi:10.1016/j.neuron.2005.06.008)
52. Harrison LM, Bestmann S, Rosa MJ, Penny W, Green GG. 2011 Time scales of representation in the human brain: weighing past information to predict future events. *Front. Hum. Neurosci.* **5**, 37. (doi:10.3389/fnhum.2011.00037)
53. Bornstein AM, Daw ND. 2012 Dissociating hippocampal and striatal contributions to sequential prediction learning. *Eur. J. Neurosci.* **35**, 1011–1023. (doi:10.1111/j.1460-9568.2011.07920.x)
54. Behrens TE, Woolrich MW, Walton ME, Rushworth MF. 2007 Learning the value of information in an uncertain world. *Nat. Neurosci.* **10**, 1214–1221. (doi:10.1038/nn1954)
55. Tobia MJ, Iacovella V, Hasson U. 2012 Multiple sensitivity profiles to diversity and transition structure in non-stationary input. *Neuroimage* **60**, 991–1005. (doi:10.1016/j.neuroimage.2012.01.041)
56. Yu AJ, Dayan P. 2005 Uncertainty, neuromodulation, and attention. *Neuron* **46**, 681–692. (doi:10.1016/j.neuron.2005.04.026)
57. Zacks JM, Braver TS, Sheridan MA, Donaldson DI, Snyder AZ, Ollinger JM, Buckner RL, Raichle ME. 2001 Human brain activity time-locked to perceptual event boundaries. *Nat. Neurosci.* **4**, 651–655. (doi:10.1038/88486)
58. Zacks JM, Speer NK, Reynolds JR. 2009 Segmentation in reading and film comprehension. *J. Exp. Psychol. Gen.* **138**, 307–327. (doi:10.1037/a0015305)
59. Tobia MJ, Iacovella V, Davis B, Hasson U. 2012 Neural systems mediating recognition of changes in statistical regularities. *Neuroimage* **63**, 1730–1742. (doi:10.1016/j.neuroimage.2012.08.017)
60. Loewenstein G. 1994 The psychology of curiosity: a review and reinterpretation. *Psychol. Bull.* **116**, 75–98. (doi:10.1037/0033-2909.116.1.75)
61. Kintsch W. 2012 Musings about beauty. *Cogn. Sci.* **36**, 635–654. (doi:10.1111/j.1551-6709.2011.01229.x)
62. Nastase SA, Iacovella V, Davis B, Hasson U. 2015 Connectivity in the human brain dissociates entropy and complexity of auditory inputs. *Neuroimage* **108**, 292–300. (doi:10.1016/j.neuroimage.2014.12.048)
63. Vitz PC. 1964 Preferences for rates of information presented by sequences of tones. *J. Exp. Psychol.* **68**, 176–183. (doi:10.1037/h0043402)
64. Shiner JS, Davison M, Landsberg PT. 1999 Simple measure for complexity. *Phys. Rev. E* **59**, 1459–1464. (doi:10.1103/PhysRevE.59.1459)
65. Hebb DO. 1949 *The organization of behavior: a neuropsychological theory*. New York, NY: John Wiley & Sons.
66. Perruchet P, Pacton S. 2006 Implicit learning and statistical learning: one phenomenon, two approaches. *Trends Cogn. Sci.* **10**, 233–238. (doi:10.1016/j.tics.2006.03.006)
67. Karuza EA, Li P, Weiss DJ, Bulgarelli F, Zinszer BD, Aslin RN. 2016 Sampling over nonuniform distributions: a neural efficiency account of the primacy effect in statistical learning. *J. Cogn. Neurosci.* **28**, 1484–1500. (doi:10.1162/jocn_a_00990)
68. Ma WJ, Jazayeri M. 2014 Neural coding of uncertainty and probability. *Annu. Rev. Neurosci.* **37**, 205–220. (doi:10.1146/annurev-neuro-071013-014017)
69. Friston K, Buzsaki G. 2016 The functional anatomy of time: what and when in the brain. *Trends Cogn. Sci.* **20**, 500–511. (doi:10.1016/j.tics.2016.05.001)
70. Dehaene S, Meyniel F, Wacongne C, Wang L, Pallier C. 2015 The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron* **88**, 2–19. (doi:10.1016/j.neuron.2015.09.019)
71. Bastos AM, Urey WM, Adams RA, Mangun GR, Fries P, Friston KJ. 2012 Canonical microcircuits for predictive coding. *Neuron* **76**, 695–711. (doi:10.1016/j.neuron.2012.10.038)
72. Davis B, Nastase S, Hasson U. In preparation. The orbitofrontal cortex is a modality-general hub for statistical learning.
73. Erickson LC, Thiessen ED. 2015 Statistical learning of language: theory, validity, and predictions of a statistical learning account of language acquisition. *Dev. Rev.* **37**, 66–108. (doi:10.1016/j.dr.2015.05.002)
74. Andric M, Hasson U. 2015 Global features of functional brain networks change with contextual disorder. *Neuroimage* **117**, 103–113. (doi:10.1016/j.neuroimage.2015.05.025)
75. Alavash M, Hilgetag CC, Thiel CM, Giessing C. 2015 Persistency and flexibility of complex brain networks underlie dual-task interference. *Hum. Brain Mapp.* **36**, 3542–3562. (doi:10.1002/hbm.22861)