Opinion piece

Author for correspondence:
Gerry T. M. Altmann
e-mail: gerry.altmann@uconn.edu

Abstraction and generalization in statistical learning: implications for the relationship between semantic types and episodic tokens

Gerry T. M. Altmann

Department of Psychological Sciences, Institute for the Brain and Cognitive Sciences, University of Connecticut, 406 Babbidge Road, Storrs, CT 06269-1020, USA

GTMA, 0000-0001-8353-1381

Statistical approaches to emergent knowledge have tended to focus on the process by which experience of *individual* episodes accumulates into generalizable experience *across* episodes. However, there is a seemingly opposite, but equally critical, process that such experience affords: the process by which, from a space of types (e.g. onions—a semantic class that develops through exposure to individual episodes involving individual onions), we can perceive or create, on-the-fly, a specific token (a specific onion, perhaps one that is chopped) in the absence of any prior perceptual experience with that specific token. This article reviews a selection of statistical learning studies that lead to the speculation that this process—the generation, on the basis of semantic memory, of a novel episodic representation—is itself an instance of a statistical, in fact associative, process. The article concludes that the same processes that enable statistical abstraction across individual episodes to form semantic memories also enable the generation, from those semantic memories, of representations that correspond to individual tokens, and of novel episodic facts about those tokens. Statistical learning is a window onto these deeper processes that underpin cognition.

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

1. Introduction

Statistical learning, as a field, has generally been concerned with mapping individual learning episodes onto abstract internal novel representations (or the labels that refer to such representations). Here, we shall focus on the reverse issue: how, from abstract (semantic) representations, we generate novel individual (episodic) representations *on-the-fly*, during language comprehension. In the context of cognition, more broadly, this corresponds to the generation of episodic 'tokens' from semantic 'types'. The latter correspond to semantic representations of typical things or the typical events that they take part in, such as the class of onions and knowledge of the events they typically participate in (henceforth, the term 'semantic memory' will be used in a narrow sense, to refer to what elsewhere is referred to as conceptual knowledge; [1]). The former are the specific examples or instances of events or things that we can perceive or create on-the-fly during language comprehension, such as the specific onion referred to in 'she peeled an onion, and chopped it finely' (cf. the episodic/semantic distinction; [2]).

The term 'statistical learning' covers a number of diverse empirical phenomena, domains of study and associated experimental paradigms. The article briefly considers the role of abstraction in statistical learning before asking: *What are the consequences of statistical learning beyond simply acquiring knowledge of the statistics of the input? Might the mechanisms underpinning statistical learning underpin the relationship between semantic types and novel episodic tokens?* The claim, building also on concepts from research on episodic memory, will be that they do.

The aim of this article is to identify the *computational principles* that underpin our ability to construct, and indeed experience, tokens on-the-fly ('tokenization'—cf. [3]). Computational and behavioural data will be reviewed that bear on the issue. It should be noted that not all of these data are from models or behavioural explorations that were originally intended to bear on episodic and/or semantic memory— sometimes, they were; other times they were not (it should be clear from the text which are which). But in the latter case, what will count are the underlying principles or properties, of the model or data, which combined together will provide a *conceptual* model (as distinct from an *implemented* model) of tokenization and the interplay between episodic and semantic memory in (certain kinds of) language comprehension.

2. Statistical learning, abstraction and the distinction between episodic and semantic knowledge

A hallmark of statistical learning is that experience of individual episodes accumulates into emergent sensitivity to the statistical properties of the input contained within each episode. Abstraction can be operationalized here as the emergence of sensitivities that (probabilistically) reflect the contexts in which each element of a stimulus occurs (and the frequency with which each such element occurs). The term 'abstraction' is often used to refer to two different, but often related, concepts: on the one hand, abstraction is the accumulation of information across individual learning episodes that gives rise to knowledge which, although not necessarily contained within any one such episode, captures (statistical) regularities across those episodes (what [4] refer to in their model of statistical learning as a process of 'integration'; see [5]). Where such regularities include co-occurrence or conditional statistics, their encoding constitutes a probabilistic encoding of the context in which a given element occurs. On the other hand, 'abstraction' is also commonly used to describe the process by which individual details of an episode are lost (potentially resulting in representations that are independent of the sensorimotor instantiation of each experience). For example, you know what a dog is, but may not recall all instances of the dogs you have encountered, nor the corresponding episodes, which nonetheless have contributed to your concept of dogs.

In some computational formalisms (e.g. 'emergentist' approaches to cognition based on interactive activation and competition; [6]), these two kinds of abstraction come about through a single mechanism and are reflected in the nature of the knowledge that is encoded within the system (see [7] for explanation). Typically, such formalisms include computations across (often distributed) representations of the input, based on the principles of associative and error-driven learning, in which input at time 1 predicts one or more inputs at time 2 or beyond in proportion to the frequency of those inputs and of their co-occurrence (cf. connectionist accounts of statistical learning; [8,9]). These formalisms are 'emergentist', because internal representations develop, or emerge, gradually through exposure to successive inputs. In other formalisms, such as the instance-based approaches to episodic and semantic memory ([10,11]; both based on the Minerva framework; [12]), the emergence of statistical regularities and the loss of detail come about through distinct mechanisms. These formalisms often involve

memorization of fragments or 'chunks' which serve as the basis for subsequent similarity matching ([13–15]; cf. [4]). Within such approaches, statistical abstraction comes about through properties of the retrieval process (which matches retrieval cues against stored fragments) rather than, as in emergentist approaches, properties of the encoding. Whereas the encoded representations in emergentist systems gradually evolve in response to subsequent input (reflecting a form of fine-tuning of the developing semantic categories), encoded representations in instance-based systems remain stable (beyond any decay mechanism; see [15] for an example of a chunking model in which the encoded chunks strengthen or decay through associative learning mechanisms).

Abstraction and generalization (applying to novel episodic content what has been learned from the training input)[1] can be related to the distinction between episodic and semantic memory. Accounts of semantic memory (narrowly defined—see above) typically assume an experiential (and often statistical) basis for semantic knowledge [16,17]. Within such accounts (putting aside, for a moment, the instance-based accounts), episodic experience (the input) is abstracted over to generate semantic knowledge that captures the statistical regularities of that experience. This semantic knowledge is the basis for subsequent generalization. But how does that process of abstraction proceed? The aim here is not to detail all possible accounts of abstraction (or to decide between the instance-based or emergentist approaches). Instead, it is to identify a critical characteristic of episodic experience that underpins both the route from episodic experience to semantic memory and, as we shall see, the reverse route also—from semantic memory to episodic tokenization. To cut the longer story short, it is the indiscriminate association of each element in an episodic context with each other element in that context (relational binding; cf. [18]). This characteristic can be deduced from any study of statistical learning (how can the cognitive system 'know' which statistical dependencies matter if it does not evaluate, or at least sample from, all possible associations?). A study by Smith & Yu [19] exemplifies this characteristic in the context of the acquisition of object–label mappings, and the study's simplicity most easily reveals the utility of indiscriminate association as a basis for abstraction and generalization.

Smith & Yu [19] asked how it could be that infants observe many objects in the environment, simultaneously hear many words potentially referring to those objects, and yet somehow work out which words should be paired with which objects. In effect, they asked how information abstracted across individual learning episodes could develop into specific object–label mappings that could be applied in (i.e. generalized to) novel episodic settings. Specifically, they asked whether knowledge of individual object–label mappings would arise from the statistics with which, across trials, multiple objects would co-occur with multiple labels. In each learning trial of their study, a pair of objects were shown to infants between 11 and 15 months of age, accompanied by a pair of words. The problem was that, on each trial, the infant could not possibly know which of the two objects should be paired with which of the two words (the 'Gavagai' problem; [20]).

Smith & Yu [19] reasoned as follows: if in the first learning trial, the two objects were a bat and a ball, and the two words were 'bat' and 'ball' (in the study, nonsense objects and nonsense words were in fact used), the ball could, in principle, become associated both with 'bat' and with 'ball', and the bat

could similarly become associated both with 'bat and 'ball'. Thus, on the basis of a single trial, the infant could not know which was the 'correct' pairing (given the potential to indiscriminately pair each object with each label). Consider, now, a second trial in which the two objects were a dog and a ball, and the two words were 'ball' and 'dog'. In this case, analogous to the first, the dog could, in principle, be associated with 'ball' and with 'dog', and the ball could also be associated with 'ball' and with 'dog'. On a purely associative learning account (there are others, but this is the most straightforward) in which statistical learning occurs across trials, and critically, in which each object would be indiscriminately associated with each label (and vice versa), the ball would be associated with the label 'ball' twice as often as any other object was associated with this same label. The strength of this ball–'ball' association would, therefore, be greater than any other association with the ball or with the word 'ball'. In the actual experiment, a task after the learning phase revealed that infants had indeed learned the cross-trial, i.e. cross-situational, statistical associations between objects and their corresponding labels.

But what has this to do with the distinction between episodic and semantic memory? In Smith & Yu [19], each trial constituted an episode, with the accumulation of statistical (co-occurrence) information leading to a form of abstraction in which the relationship between one particular label and one particular object *across trials* presumably became more salient than their relationships to other labels and the objects with which they incidentally co-occurred on individual trials. Abstraction can be operationalized here as an accumulation of experience across trials which leads to generalizable knowledge (the 'correct' object–label mappings) not available to the organism at the start of this accumulation, and in which episode-specific details (including non-systematic, accidental, co-occurrences, including the 'incorrect' object–label pairings as well as the object–object and label–label pairings) become less salient than more systematic details, reflecting structure (or regularities) across episodes. Moreover, to the extent that object–label mappings are a part of semantic memory (the mappings, once acquired, are relatively constant, regardless of the episodic contexts in which the objects and/or their corresponding labels occur—hence 'semantic'), the transition in this study from individual episodes to semantic memory came about through a process of statistical abstraction, with statistical regularities emerging across successive trials.

In the following, we briefly consider a variety of computational models of abstraction that embody this same indiscriminate association of each 'thing' in an episodic context with each other thing. Our theoretical treatment of tokenization, which can be thought of as the 'reverse' of abstraction (insofar as novel episodic representations are generated from pre-established semantic representations), will rely heavily on this same principle.

3. From episodic to semantic memory: (some) computational principles

There exist a variety of computational and statistical models of memory which embody principles relevant to abstraction and the episodic/semantic distinction. In Lund & Burgess's [21] hyperspace analogue to language (HAL), a semantic space was generated by calculating co-occurrences between all words in a language corpus (within a sliding window of fixed width) and treating the ensuing matrix as a high-dimensional space that was then reduced using principal components analysis to generate a semantic space in which proximity within the space corresponded to semantic similarity. The contents of the sliding window constituted the episodic context of each word it contained, and the calculation of co-occurrence within the window was a largely indiscriminate process that associated each word with each other word. The principles of HAL are not unlike those described in Elman's work with simple recurrent networks (SRN: [22]): Elman's network learned aspects of the sequential co-occurrence statistics presented to it in short language-like sequences. It did so by having to predict at each moment in time what the input would be at the next moment in time. The discrepancy between its prediction and the actual input at that next moment drove changes to the weights on its internal connections. The result of this error-driven learning was the emergent encoding of a similarity space in which words referring to similar objects or actions were located more closely to one another than to words referring to dissimilar objects or actions (see also [16]). Both HAL and the SRN produced reasonable similarity spaces because dependencies in language map onto dependencies in the real world: e.g. the kinds of words that occur after a verb such as 'eat' are going to refer to the kinds of objects that in the real world can be eaten; thus, capturing their dependency in language captures the semantic similarity between their real-world referents. Other corpus-based approaches (e.g. BEAGLE; [23]; LSA; [24]) capture such semantic similarities in similar ways (see [25] for a brief review of distributional models of semantic memory). In these cases, the dimensional reduction of the corpus statistics to generate a semantic similarity space constitutes the abstraction from individual episodes (of co-occurrence) to generalizable experience encoded in a multidimensional semantic space (semantic memory).

Altmann & Mirkovic [26] described how a modified version of the SRN [27–29] might support the mapping from unfolding language to event representation (see also [30–32]). We claimed that this mapping manifests as the ability to predict both how the language will unfold, and how the real-world event described by that language would unfold if it were being experienced directly.[2] Whereas Elman's network consisted of a single set of input units, hidden (and copy-through-time or recurrent) units, and output units, the modified SRN contained two sets of input units (one for each domain of input), two layers of hidden units (the second connected to a recurrent layer that fed back into this second layer) and two sets of output units (again, one for each domain). Both sets of input units fed into the same (first) hidden layer. We argued that the representational substrate (the hidden layers) common to both the linguistic and non-linguistic domains allowed the predictive process to operate over variable time frames and variable levels of representational abstraction. Here, like before, 'abstraction' refers to the emergence of sensitivity to predictive contingencies (i.e. distributional statistics) through time, but importantly, it also includes sensitivity to contingencies through levels of (emergent) *hierarchical* representation (in which higher-level emergent representations can be composed of lower-level representations that vary across shorter time frames; cf. words-to-syllables-to-phonemes).

The relevance to the episodic/semantic distinction is that the emergent 'representations' of the SRN are akin to

semantic memory—they can be conceived as long-term, potentially hierarchically organized, knowledge abstracted across individual episodes of input (encoded in the weights on the connections between individual units within the networks). Episodic memory in the SRN is a more complex matter: input to the SRN feeds to the hidden layer(s) whose input is not simply this current input, but also a copy of its own immediately prior activation state. This prior activation state encodes not only the 'echoes' of prior states (and their concurrent inputs), but also longer-term semantic knowledge that emerged through gradual changes in the network's internal connection strengths (as determined by the goodness-of-fit between the network's output and, for Elman's prediction task, the next input). Episodic and semantic knowledge are thus combined through time, with no clear separation between the two. Short-term changes in connectivity essentially encode the episodic context superimposed on both the prior context and the longer-term encoding of semantic memory, itself gradually evolving in response to those individual episodes. But the critical principle here, beyond the relationship between emergence and abstraction in the SRN (which relies, again, on initially indiscriminate associations; see [7] for discussion), is the idea, enshrined in recurrence through time, that the current input and its concurrent (episodic) context can become combined, or associated, not simply at a single moment in time, but *across time* with prior inputs and prior episodic contexts. Moreover, modulated by the network's internal connectivity, everything in the current input can potentially be associated with everything else in the current input or elsewhere in the network's representational space (a by-product of spreading activation across time). Whether those associations have any impact on that representational space (i.e. become more lastingly associated with it or change it) depends on their statistical relationship to the information encoded within that part of the space. As a mechanism, this has the ingredients required to support the transition, assumed to underpin the Smith & Yu [19] data, from initially indiscriminate associations to subsequent systematic cross-domain mapping.

The SRN is not, and was never intended to be, a model of the relationship between human episodic and semantic memory. Such models, typically based on complementary learning systems (CLS: [36]), postulate a clearer distinction between the two than is manifest in the SRN. Unlike the SRN, these models tend to take inspiration from the neurobiology of the neocortical and hippocampal structures in the brain that support memory ([36–40]; see [41] for a non-connectionist model of episodic and semantic memory that is also based on hippocampal function). A basic tenet of CLS is that hippocampal structures support the rapid encoding of distinct episodes (i.e. episodic memory) through large changes in connectivity both within hippocampus and between hippocampus and neocortex, whereas smaller changes in connectivity within neocortex support slow encoding of regularities encountered across multiple episodes (i.e. semantic memory). The two systems are also complementary in the sense that neocortex captures similarity through pattern overlap, whereas hippocampus maintains the distinctiveness of episodes by encoding them as more separate, sparse, patterns. Recently, Kumaran & McClelland [38,40] have argued that 'big loop' recurrence between areas CA3 and CA1 (within 'hippocampus proper') and entorhinal cortex (essentially, the input to, and output from, hippocampus; the 'interface' to neocortex) is an essential architectural ingredient for several of the hippocampus's abilities. These include (among others) its ability to discover higher-order (i.e. abstract) structure across episodes; to generalize (e.g. to form novel associations between items that were not co-present—see also [40]); to encode individual experience as arbitrary combinations of elements; and to arbitrarily recombine elements in memory during 'constructive' memory (cf. [42,43]). Interestingly, Kumaran & McClelland [38] implemented their model of hippocampal function in an architecture that has properties of both exemplar/instance-based models *and* connectionist (error-driven associative learning) models.

In the following, we revisit many of the phenomena and principles described above, but the focus now shifts to consideration of their application to the challenge of tokenization.

4. From semantic memory to episodic realization

The studies described thus far, like the majority of studies on statistical learning (broadly construed), focus on the route from input (i.e. episodes) to abstraction, i.e. the experiential basis for abstract representations that can support generalization of that prior experience to novel contexts. However, there is one aspect of language comprehension that presents a particular challenge to classical accounts of episodic/semantic memory and abstract representation: we can use language not simply to refer to an event that did occur, in which case language serves as a cue to the retrieval from memory of the appropriate episode, but to refer to an event that did not, as in 'the woman chopped an onion. Then, she fried it' (assuming that this does not correspond to any specific memory the reader can recall). In this case, we instantiate in the first sentence (and hence, in our mental representation of the corresponding possible world) a woman, an onion and changes in the state of the onion (for discussion of object state change, see [44–46]). In the second sentence, we refer back to that onion in the context of its chopped 'version' being fried. Thus, when comprehending the first sentence and the corresponding event, we need to access semantic memory to instantiate *as episodic entities* bound in some temporary timeframe the woman and the onion (at a minimum). How do we do this?

Moreover, we have to bind the distinct states of the onion into distinct temporally separated episodes (first the episode that transforms its state from intact to chopped, and then the episode that transforms it from raw to fried)[3]. Thus, whereas language *learning* (and experience more generally) may require the route from episodic experience to semantic representation (via abstraction), language *comprehension* requires the route from semantic to episodic representation. Whereas the literature on episodic memory (see [48] for review from a neuroscientific perspective) has focused on episodic representation as it applies to the encoding and recollection of experienced events, here we focus on episodic representation insofar as language can evoke novel episodic instances, as in the chopping example above ([43] refer to such evocation as 'construction' or 'scene construction'). In this regard, language is able to take the comprehender beyond the confines of memory.

Can insights from statistical learning inform accounts of this route from semantic memory to novel instantiation of episodic tokens? Studies with infants and with adults demonstrate

5

that statistical learning can occur in the absence of any instruction to extract statistics or look for patterns (within or across sensory modalities), and on the basis that, on any given trial or exposure, it is not possible to know which dependencies/ patterns are critical and which are incidental; the learner must, in effect, associate everything with everything else, with the strength of each association changing, with continued experience, to become a reflection of its utility in respect of predicting future associations. This is the associative learning mechanism postulated by Smith & Yu [19] to explain their results. This kind of indiscriminate association underpins accounts of 'relational memory' and relational binding in episodic memory ([18]; see e.g. [49] for review). As we shall now see, it also underpins tokenization, whether through direct sensory experience, memory retrieval, or language comprehension.

The challenge of tokenization, as set out above, is to understand how, on hearing 'The woman chopped an onion. Then, she fried it', we can instantiate novel episodic representations corresponding to a woman and an onion in its different states (intact, chopped, fried) even if we have never experienced this particular woman and this particular onion first-hand. But, in fact, the same instantiations of these entities and their corresponding states have to be created if we directly experience a woman (whom we may or may not know) chopping an onion and then frying it: the episodic experience of the woman, in this case, consists of perceptual features which are bound both to each other and to semantic features (semantic memory) via hippocampal–cortical interactions (see e.g. [48] for review). The perceptual features are not just those of the woman, but of the spatial context, and other elements also in that context. In effect, discrete sets of perceptual features are associated with one another despite the arbitrariness of their co-occurrence. So in terms of direct (i.e. real-world) experience, the woman becomes associated with incidental features of her context in much the same way as the visual stimuli presented to the infants in the Smith & Yu [19] study became associated with whatever co-occurred with them (in that case, the two spoken labels).

The perceptual features associated with the perceived woman activate (and by association, bind to) the semantic memory associated both with this particular woman (if known to the viewer) and with women in general (although as outlined in [50], only the contextually relevant details of such general information will probably be activated). So, on the one hand, tokenization relies on *indiscriminate* association of perceptual (or other) features with other elements in those features' (sensorimotoric) contexts; and on the other hand, it relies on more *systematic* associations between those features and representations in semantic memory.[4] However, there is a third critical ingredient also: these distinct sets of associations have also to be *grounded in time*. This can be achieved if we assume that the current context is accompanied by echoes of contexts past, as afforded by recurrence through time (see above for the discussion of models of hippocampal function that include such recurrence). Recurrence through time ensures that elements in the current context (i.e. discrete sets of features, whether or not perceptually grounded) will associate not simply with other elements in the shared context, but with echoes of those (and other) elements from previous contexts. A corollary of such recurrence is that as time moves on, what had been the current context becomes itself an echo with which elements in its future can also

associate. Conceptually, this is similar to the activation dynamics across time seen in the SRN; the activation state of the network reflects the modulation of activation owing to the current input by activation owing to the prior inputs. This, in turn, means that changes in connectivity at any moment in time are due to the interactions between the input at that time and the input at past times; in effect, these interactions bind the network's internal representations at one moment in time with its internal representations at prior moments in time.

When we directly perceive the onion, the account of how it becomes a 'token' (i.e. an episodic instance of a semantic category) is the same: we associate it with other things in the concurrent context, and with these and other things from prior, spatio-temporally contiguous, contexts. That is, we ground it in space (defined through location relative to other objects in the context) and time. These associations are partly systematic (to the extent that there may be predictive contingencies between e.g. onions and chopping boards which are consistent with past experience) and partly non-systematic (i.e. accidental co-occurrences between the elements of the concurrent and past contexts; e.g. the onion and the cat that had slouched out of the room). In addition, our episodic experience of the onion includes the systematic associations between the onion's perceptual features and our abstract knowledge of onions as a class of thing (i.e. the semantic memory of what onions *are*). When we see the woman fry the onion after it has been chopped, spatio-temporal continuity across the different states of the onion allows us to infer (to use the term informally) that what is being fried is episodically related to what had been chopped (i.e. the episodic experience of the intact onion is bound through spatio-temporal continuity to the episodic experience of the chopped onion and, subsequently, to that of the onion being fried; e.g. [53])[5].

However, what if there has been no such direct experience when hearing *The woman chopped an onion. Then, she fried it*? In this case also, just like with the direct experience of the actual real-world event, there is an episodic experience (i.e. bound in space and time) comprising a set of perceptual features bound both to each other and to semantic memory via, one might suppose, similar hippocampal–cortical interactions. The perceptual features associated with the phrase 'the woman' do not just associate with one another and with the semantic knowledge corresponding to the meaning of the phrase; they also associate with the incidental features of the accompanying linguistic and non-linguistic context—the location, time and other incidental features co-occurring in contiguous space–time with the experience of that phrase (whether spoken or written). Similarly, for the onion. The associations with the current context create the phenomenological experience of the word (just as in the directly experienced analogue, these associations provide the basis for the conscious experience of the woman or the onion). The ensemble of associations that include the perceptual/semantic characteristics elicited by each word, and the co-occurrences with other features of the current contexts (and past echoes) within which the word is perceived, constitute the constructed 'token' of the woman, or the onion. In terms of mechanism, this is just the same as when actually seeing a woman in a context 'tokenizes' her.

Earlier, we distinguished between a route from episodic experience to semantic memory and the converse route, from semantic representation to episodic realization. It turns out,

however, that the routes are not so different after all: in terms of the phenomenology of (episodic) experience, and specifically the experience of objects as instances of things, and of words as instances of reference to things, the route travelled is essentially the same. Relational binding with the context, essentially a fast-mapping (i.e. one trial) associative process, is the basis for the distinction between types (as encoded in semantic memory) and tokens (the products of episodic experience).

5. Conclusions: a role for statistical learning in understanding memory systems

Where we have ended up, discussing types, tokens and the phenomenology of experience, may at first blush appear a far cry from our earlier discussion of the cross-situational disambiguation of possible word–label pairings. However, the associative processes that underpin statistical learning are not so dissimilar from the processes that may control the episodic–semantic interface whether during language processing or cognition more generally. These processes include the relational binding of the elements in immediate experience; the mapping of associative relations in semantic memory to these associative relations in immediate experience; the abstraction across experience as the stripping away of non-systematic associative relations (cf. the representations that 'emerge' in Elman's SRN); and the ability to apply that experience to novel situations with novel relations (i.e. to generalize). Equally, there are aspects of (some kinds of) statistical learning that appear less critical; sequential structure in statistical learning paradigms may be important for its parallels with structure in spoken language (or with sequential structure in the visual domain) but may be 'just another' example of structure amenable to error-driven associative learning. The parallels between processes of statistical learning and those underpinning the relationship between episodic and semantic memory suggest that accidental co-occurrences (whether in or across time) are as important to statistical learning as are systematic co-occurrences. They are not 'noise' but are critical not just because, initially, the cognitive system cannot know what is systematic and what is accidental, but because, subsequently, this indiscriminate association of one thing with another is the very basis for our ability to distinguish between different instances of the same kind of object, or between one instance of an object and another instance of that same object.

Theoretically, the account of tokenization developed above, based on relational binding, offers a novel perspective on action and event representation, as well as some challenges. If a chopped onion is associated with its prior intact self, through relational binding across time, then it carries with it its own 'history'—of its states and changes to its states across time (cf. 'perdurance' theories of object persistence; see [53,54] for review). This creates a challenge: when selecting a representation of that onion, multiple representations of that onion in different states may be available, but only the situationally appropriate one must be chosen (e.g. when referring to that onion before it was chopped). Hence, referring to an object when it has undergone a change in state requires a competitive process that selects from among multiple possible alternatives (selecting one at the expense of others). We have found across a number of studies precisely such competition [44–46]. Moreover, we observe competition only between

different states of the *same* object token; we do not find competition between different object states when they are associated with *different* tokens [46]. There are theoretical consequences also of an object carrying with it its own history: What is an event if not the spatio-temporally intersecting histories of the objects that participate in that event? If events are indeed represented through intersecting object representations (given that these represent their changing states), *actions* can be considered emergent representations abstracted across those intersecting object representations [55]. As such, actions are not foundational primitives of event representation.

The account developed above also allows for an interesting prediction. It relies on the same mechanisms that underpin the relational memory approach to episodic recollection and hippocampal function [18], namely those mechanisms that enable the formation of arbitrary associations between the elements within an individual experience. These are typically assumed to be hippocampal in origin [48,56]. The prediction, then, is that damage to hippocampus should impair the ability to recognize entities as *tokens*, and specifically to track changes to those tokens across time. Evidence from patients with hippocampal damage suggests that, in the domain of language at least, there *are* impairments in the ability to refer to individual tokens, both in respect of the processes that enable referential continuity and coherence from one sentence to another [57] and in respect of the processing of pronouns such as 'it' or 'she' which refer to specific individual tokens [58]. However, further research is needed to establish whether hippocampal damage also impairs the ability to recognize that, for example, a chopped onion is the *same* onion as the one that was intact beforehand. It is unclear to what extent such an impairment would manifest in day-to-day function: typically, not very much would be consequential on whether it was indeed the same onion, or a different one, and for this reason, such deficits, if they do occur, may not be so obvious in patients with hippocampal damage. Importantly, whereas tokenization relies on relational binding, other aspects of comprehension do not; they rely instead on non-arbitrary binding to pre-existing knowledge and are supported by a variety of distinct brain regions other than the hippocampal complex. For language comprehension, these include regions most typically associated with the 'language network' (e.g. inferior frontal gyrus, the superior and middle temporal gyri, and the angular gyrus; [59]) as well as other regions associated with the integration of semantic knowledge, such as medial prefrontal cortex (cf. [60]) and perirhinal cortex (cf. [61]). These other kinds of binding—in essence, the non-arbitrary aspects of language processing—enable essentially intact language comprehension in the face of severe hippocampal damage.

Given the role of the hippocampus in associative learning, in relational and episodic memory, and as an interface to semantic memory, theories of hippocampal function will increasingly inform accounts of statistical learning (cf. [40,62,63]). In this respect, it is relevant that the hippocampus is not a homogeneous structure, but contains substructures that appear to support a 'gradient of abstraction', with more perceptually grounded, spatio-temporally fine-grained, representational properties in more posterior parts and less grounded, coarser-grained, and more abstract representation towards more anterior parts [64,65]. This suggests that more detailed analyses of hippocampal activity could even be diagnostic of the kinds of emergent abstraction that develop during statistical/error-driven associative learning, as well as of the

kinds of representation that are constructed on-the-fly during language processing.

Taking a broader perspective on statistical learning allows more detailed probing of cognitive function; not from the perspective of trying to understand how language may be acquired from individual learning episodes, or visual input made sense of (cf. visual statistical learning), but from the perspective of trying to understand the relationship between the different learning and memory systems that underpin experience and memory, and abstraction and generalization (see [66,67] for an example of this broader perspective, in which sleep takes on a central role in statistical learning). The distinction between semantic types and semantic tokens is just one small component in a broader theoretical landscape populated by a wealth of relevant behavioural, neuroscientific and computational data. The challenge is to abstract across it all.

Endnotes

[1]Henceforth, 'generalization' will be used to denote the *application* of knowledge in novel situations, often manifesting as overt behaviour. Often, the term is used to denote what is, in fact, a process of abstraction (e.g. 'generalizing across instances to infer a common characteristic'), but for the sake of consistency we limit generalization here to the application of knowledge (in novel situational or episodic contexts), and abstraction to a form of induction (across experience) of new knowledge.

[2]The model had originally been developed to explain cross-domain transfer effects in artificial grammar learning [33]. In these cases, participants exposed to sequences of *auditory* monosyllabic non-words were able to classify sequences of novel arbitrary *visual* (graphical) symbols according to whether they obeyed the same 'rules' that generated the syllables they had heard previously. Tunney & Altmann [34] demonstrated that there are two dissociable mechanisms that afford such cross-domain transfer of knowledge: one based on abstract analogy of 'repetition structure' [35], in which a sequence such as 'jix pel jix sog' can be mapped onto '+ *+=' (novel symbols were used, not the mathematical symbols shown here), and importantly, another based solely on computing, and mapping between, statistical distributions of non-repeating elements (e.g. 'jix pel het sog' onto '+ *~=' – unlike abstract analogy, this process requires statistical abstraction across multiple exemplars to induce the statistical distributions). The model was developed specifically to explain this latter case, i.e. the mapping of statistical patterns in one domain, onto statistical patterns in the other.

[3]It is unclear whether these episodes are, in fact, 'temporally separated'. There is, necessarily, continuity from one episode to the next, and most likely their representations overlap. Temporal discontinuity may manifest in their encoding as discontinuities in predictability of episodic state across time (cf. the computational instantiation of event segmentation theory described in [47]).

[4]The current account of tokenization differs from that developed in the context of visual perception by Bowman & Wyble [3]; they used the term 'tokenization' to refer to the process by which 'types' of perceptual feature (e.g. colour and orientation) are bound together to create perceptual 'tokens' (visual objects)—cf. Zimmer and Ecker [51]. Theirs is a model of visual attention, (visual) working memory, and the attentional blink [52], and is not intended to address the semantic type—episodic token distinction as used here.

[5]In fact, spatio-temporal continuity is not required to explain object persistence across change: if the transition from intact to chopped is occluded, the onion in its chopped state will activate semantic knowledge of onions, in general, which will re-activate the episodic memory of the previously seen intact onion (its recency gives it prepotency in respect of its activation state). This latter representation will, by virtue of its co-activation with the currently seen chopped onion, become associated through time with the chopped onion. This form of semantically mediated associative/relational binding may be sufficient to support the experience of object persistence across changes in which the distinct states of the object are each recognizable as belonging to the same semantic type. More generally, semantic mediation enables non-spatiotemporally contiguous elements to become a part of the 'context' associated with an individual episodic token.

References

1. Yee E, Chrysikou EG, Thompson-Schill SL. 2013 Semantic Memory. In *The Oxford Handbook of Cognitive Neuroscience* (eds K Ochsner, S Kosslyn), pp. 353–374. Oxford: Oxford University Press.

2. Tulving E. 1972 Episodic and semantic memory. In *Organization of memory* (eds E Tulving, W Donaldson), pp. 382–403. Academic Press.

3. Bowman H, Wyble B. 2007 The simultaneous type, serial token model of temporal attention and working memory. *Psychol. Rev.* **114**, 38–70. (doi:10.1037/0033-295X.114.1.38)

4. 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)

5. Thiessen ED. 2017 What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Phil. Trans. R. Soc. B* **372**, 20160056. (doi:10.1098/rstb.2016.0056)

6. Elman JL, Bates EA, Johnson MH, Karmiloff-Smith A, Parisi D, Plunkett K. 1996 *Rethinking innateness*. Cambridge, MA: MIT Press.

7. Altmann GT. 1997 *The ascent of babel: an exploration of language. Mind and understanding*. Oxford, UK: Oxford University Press.

8. Cleeremans A. 1993 *Mechanisms of implicit learning: connectionist models of sequence processing*. Cambridge, MA: MIT Press.

9. Cleeremans A, Dienes Z. 2008 Computational models of implicit learning. *Camb. Handb. Comput. Psychol.* 396–421. (doi:10.1017/CBO9780511816772.018)

10. 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)

11. Johns BT, Jones MN. 2015 Generating structure from experience: a retrieval-based model of language processing. *Can. J. Exp. Psychol.* **69**, 233–251. (doi:10.1037/cep0000053)

12. Hintzman DL. 1984 MINERVA 2: a simulation model of human memory. *Behav. Res. Methods, Instrum. Comput.* **16**, 96–101. (doi:10.3758/BF03202365)

13. Perruchet P, Pacteau C. 1990 Synthetic grammar learning: implicit rule abstraction or explicit fragmentary knowledge? *J. Exp. Psychol. Gen.* **119**, 264–275. (doi:10.1037/0096-3445.119.3.264)

14. 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)

15. Perruchet P, Vinter A. 1998 PARSER: A model for word segmentation. *J. Mem. Lang.* **39**, 246–263. (doi:10.1006/jmla.1998.2576)

16. Rogers TT, McClelland JL. 2004 *Semantic cognition: a parallel distributed processing approach*. Cambridge, MA: MIT Press.

17. Sloutsky VM. 2010 From perceptual categories to concepts: what develops? *Cogn. Sci.* **34**, 1244–1286. (doi:10.1111/j.1551-6709.2010.01129.x)

18. Cohen NJ, Eichenbaum H. 1993 *Memory, amnesia, and the hippocampal system*. Cambridge, MA: MIT Press.

19. Smith L, Yu C. 2008 Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* **106**, 1558–1568. (doi:10.1016/j.cognition.2007.06.010)

20. Quine WVO. 1960 *Word and object*. Cambridge, MA: MIT Press.

21. Lund K, Burgess C. 1996 Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods, Instrum. Comp.* **28**, 203–208. (doi:10.3758/BF03204766)

22. Elman JL. 1990 Finding structure in time. *Cogn. Sci.* **14**, 179–211. (doi:10.1207/s15516709cog1402_1)

23. Jones MN, Mewhort DJ. 2007 Representing word meaning and order information in a composite holographic lexicon. *Psychol. Rev.* **114**, 1–37. (doi:10.1037/0033-295X.114.1.1)

24. Landauer TK, Dumais ST. 1997 A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* **104**, 211–240. (doi:10.1037/0033-295X.104.2.211)

25. Jones MN, Willits JA, Dennis S. 2014 Models of semantic memory. In *Oxford handbook of mathematical and computational psychology* (eds JR Busemeyer, JT Townsend). Oxford, UK: Oxford University Press.

26. Altmann GTM, Mirkovic J. 2009 Incrementality and prediction in human sentence processing. *Cogn. Sci.* **33**, 583–609. (doi:10.1111/j.1551-6709.2009.01022.x)

27. Altmann GTM. 2002 Learning and development in neural networks: the importance of prior experience. *Cognition* **85**, 43–50. (doi:10.1016/S0010-0277(02)00106-3)

28. Altmann GTM, Dienes Z. 1999 Rule learning by seven-month-old infants and neural networks. *Science* **284**, 875. (doi:10.1126/science.284.5416.875a)

29. Dienes Z, Altmann GTM, Gao S-J. 1999 Mapping across domains without feedback: A neural network model of transfer of implicit knowledge. *Cogn. Sci.* **23**, 53–82. (doi:10.1207/s15516709cog2301_3)

30. McClelland JL, St. John M, Taraban R. 1989 Sentence comprehension: a parallel distributed processing approach. *Lang. Cogn. Process.* **4**, SI287–SI335. (doi:10.1080/01690968908406371)

31. John MF. 1992 The story Gestalt: a model of knowledge-intensive processes in text comprehension. *Cogn. Sci.* **16**, 271–306. (doi:10.1207/s15516709cog1602_5)

32. John MF, McClelland JL. 1990 Learning and applying contextual constraints in sentence comprehension. *Artif. Intell.* **46**, 217–257. (doi:10.1016/0004-3702(90)90008-N)

33. Altmann GTM, Dienes Z, Goode A. 1995 On the modality-independence of implicitly learned grammatical knowledge. *J. Exp. Psychol. Learn. Mem. Cogn.* **21**, 899–912. (doi:10.1037/0278-7393.21.4.899)

34. Tunney RJ, Altmann GTM. 2001 Two modes of transfer in artificial grammar learning. *J. Exp. Psychol. Learn. Mem. Cogn.* **27**, 614–639. (doi:10.1037/0278-7393.27.3.614)

35. Brooks LR, Vokey JR. 1991 Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews *et al.* (1989). *J. Exp. Psychol. Gen.* **120**, 316–323. (doi:10.1037/0096-3445.120.3.316)

36. McClelland JL, McNaughton BL, O'Reilly RC. 1995 Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* **102**, 419. (doi:10.1037/0033-295X.102.3.419)

37. Kumaran D, Hassabis D, McClelland JL. 2016 What learning systems do intelligent agents need? complementary learning systems theory updated. *Trends Cogn. Sci.* **20**, 512–534. (doi:10.1016/j.tics.2016.05.004)

38. Kumaran D, McClelland JL. 2012 Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. *Psychol. Rev.* **119**, 573–616. (doi:10.1037/a0028681)

39. Norman KA, O'Reilly RC. 2003 Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychol. Rev.* **110**, 611–646. (doi:10.1037/0033-295X.110.4.611)

40. Schapiro AC, Turk-Browne NB, Botvinick MM, Norman KA. 2017 Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Phil. Trans. R. Soc. B* **372**, 20160049. (doi:10.1098/rstb.2016.0049)

41. Howard MW, Shankar KH, Jagadisan UK. 2011 Constructing semantic representations from a gradually changing representation of temporal context. *Top. Cogn. Sci.* **3**, 48–73. (doi:10.1111/j.1756-8765.2010.01112.x)

42. Hassabis D, Maguire EA. 2007 Deconstructing episodic memory with construction. *Trends Cogn. Sci.* **11**, 299–306. (doi:10.1016/j.tics.2007.05.001)

43. Hassabis D, Maguire EA. 2009 The construction system of the brain. *Phil. Trans. R. Soc. B* **364**, 1263–1271. (doi:10.1098/rstb.2008.0296)

44. Hindy NC, Altmann GTM, Kalenik E, Thompson-Schill SL. 2012 The effect of object state-changes on event processing: do objects compete with themselves? *J. Neurosci.* **32**, 5795–5803. (doi:10.1523/JNEUROSCI.6294-11.2012)

45. Hindy NC, Solomon SH, Altmann GTM, Thompson-Schill SL. 2015 A cortical network for the encoding of object change. *Cereb. Cortex* **25**, 884–894. (doi:10.1093/cercor/bht275)

46. Solomon SH, Hindy NC, Altmann GT, Thompson-Schill SL. 2015 Competition between mutually exclusive object states in event comprehension. *J. Cogn. Neurosci.* **27**, 2324–2338. (doi:10.1162/jocn_a_00866)

47. Reynolds JR, Zacks JM, Braver TS. 2007 A computational model of event segmentation from perceptual prediction. *Cogn. Sci.* **31**, 613–643. (doi:10.1080/15326900701399913)

48. Moscovitch M, Cabeza R, Winocur G, Nadel L. 2016 Episodic memory and beyond: the hippocampus and neocortex in transformation. *Annu. Rev. Psychol.* **67**, 105–134. (doi:10.1146/annurev-psych-113011-143733)

49. Konkel A, Cohen NJ. 2009 Relational memory and the hippocampus: representations and methods. *Front. Neurosci.* **3**, 23. (doi:10.3389/neuro.01.023.2009)

50. Yee E, Thompson-Schill SL. 2016 Putting concepts into context. *Psychon. Bull. Rev.* **23**, 1015–1027. (doi:10.3758/s13423-015-0948-7)

51. Zimmer HD, Ecker UKH. 2010 Remembering perceptual features unequally bound in object and episodic tokens: neural mechanisms and their electrophysiological correlates. *Neurosci. Biobehav. Rev.* **34**, 1066–1079. (doi:10.1016/j.neubiorev.2010.01.014)

52. Raymond JE, Shapiro KL, Arnell KM. 1992 Temporary suppression of visual processing in an RSVP task: an attentional blink? *J. Exp. Psychol. Hum. Percept. Perform.* **18**, 849–860. (doi:10.1037/0096-1523.18.3.849)

53. Scholl BJ. 2007 Object persistence in philosophy and psychology. *Mind Lang.* **22**, 563–591. (doi:10.1111/j.1468-0017.2007.00321.x)

54. Sider T. 2001 *Four-dimensionalism*. Oxford, UK: Clarendon Press.

55. Altmann GTM, Ekves Z. In preparation. Multiple object-states represent events: A theory of event representation.

56. Konkel A, Warren DE, Duff MC, Tranel D, Cohen NJ. 2008 Hippocampal amnesia impairs all manner of relational memory. *Front. Hum. Neurosci.* **2**, 1–15. (doi:10.3389/neuro.09.015.2008)

57. Kurczek J, Duff MC. 2011 Cohesion, coherence, and declarative memory: discourse patterns in individuals with hippocampal amnesia. *Aphasiology* **25**, 700–712. (doi:10.1080/02687038.2010.537345)

58. Kurczek J, Brown-Schmidt S, Duff M. 2013 Hippocampal contributions to language: evidence of referential processing deficits in amnesia. *J. Exp. Psychol. Gen.* **142**, 1346–1354. (doi:10.1037/a0034026)

59. Friederici AD. 2011 The brain basis of language processing: from structure to function. *Physiol. Rev.* **91**, 1357–1392. (doi:10.1152/physrev.00006.2011)

60. van Kesteren MT, Ruiter DJ, Fernández G, Henson RN. 2012 How schema and novelty augment memory formation. *Trends Neurosci.* **35**, 211–219. (doi:10.1016/j.tins.2012.02.001)

61. Taylor KI, Moss HE, Stamatakis EA, Tyler LK. 2006 Binding crossmodal object features in perirhinal cortex. *Proc. Natl Acad. Sci. USA* **103**, 8239–8244. (doi:10.1073/pnas. 0509704103)

62. Gómez RL. 2017 Do infants retain the statistics of a statistical learning experience? Insights from a developmental cognitive neuroscience perspective. *Phil. Trans. R. Soc. B* **372**, 20160054. (doi:10.1098/ rstb.2016.0054)

63. Schapiro AC, Turk-Browne NB, Norman KA, Botvinick MM. 2016 Statistical learning of temporal community structure in the hippocampus. *Hippocampus* **26**, 3–8. (doi:10.1002/ hipo.22523)

64. Poppenk J, Evensmoen HR, Moscovitch M, Nadel L. 2013 Long-axis specialization of the human hippocampus. *Trends Cogn. Sci.* **17**, 230–240. (doi:10.1016/j.tics. 2013.03.005)

65. Long LL, Bunce JG, Chrobak JJ. 2015 Theta variation and spatiotemporal scaling along the septotemporal axis of the hippocampus. *Front. Syst. Neurosci.* **9**, 37. (doi:10.3389/fnsys.2015.00037)

66. Gómez RL, Bootzin RR, Nadel L. 2006 Naps promote abstraction in language-learning infants. *Psychol. Sci.* **17**, 670–674. (doi:10.1111/j.1467-9280.2006.01764.x)

67. Hupbach A, Gomez RL, Bootzin RR, Nadel L. 2009 Nap-dependent learning in infants. *Dev. Sci.* **12**, 1007–1012. (doi:10.1111/j.1467-7687.2009.00837.x)

9

rstb.royalsocietypublishing.org *Phil. Trans. R. Soc. B* **372**: 20160060