



Implicit learning and statistical learning: one phenomenon, two approaches

Pierre Perruchet and Sebastien Pacton

Université de Bourgogne, LEAD/CNRS, Pôle AAFE, Esplanade Erasme, 21000 Dijon, France

The domain-general learning mechanisms elicited in incidental learning situations are of potential interest in many research fields, including language acquisition, object knowledge formation and motor learning. They have been the focus of studies on implicit learning for nearly 40 years. Stemming from a different research tradition, studies on statistical learning carried out in the past 10 years after the seminal studies by Saffran and collaborators, appear to be closely related, and the similarity between the two approaches is strengthened further by their recent evolution. However, implicit learning and statistical learning research favor different interpretations, focusing on the formation of chunks and statistical computations, respectively. We examine these differing approaches and suggest that this divergence opens up a major theoretical challenge for future studies.

Introduction

There is no doubt that many of our most fundamental abilities, whether they concern language, perception, motor skill, or social behavior, reflect some kind of adaptation to the regularities of the world that evolves without intention to learn, and without a clear awareness of what we know. This ubiquitous phenomenon was called ‘implicit learning’ (IL) by Reber [1,2] 40 years ago. Since then, several studies have explored this form of learning with several experimental paradigms (mainly finite-state grammars and serial reaction time tasks; for reviews, see [3,4]).

Originating from a different research tradition, the term ‘statistical learning’ (SL) was proposed 10 years ago by Saffran and collaborators [4] to designate the ability of infants to discover the words embedded in a continuous artificial language, and this field of research is now growing exponentially. There are obvious similarities between SL and IL. As in IL, participants in SL experiments are faced with structured material without being instructed to learn. They learn merely from exposure to positive instances, without engaging in analytical processes or hypothesis-testing strategies. Researchers have pointed out that SL proceeds

automatically [5–8], incidentally [9], spontaneously [6], or by simple observation [9], and that participants in SL settings were unaware of the statistical structure of the material [7].

This article first describes how recent evolution in IL and SL research fields has made them closer to one another, leading to a growing number of cross-references and to the occasional use of the two expressions as synonymous. Conway and Christiansen [10] even now propose the term ‘implicit statistical learning’ to cover the two domains. However, we then go on to show that beyond the similarity of paradigms and results, the two domains emphasize different interpretations of the data. We suggest that this divergence, which has not been highlighted as yet, opens up a deep challenge for future studies.

The recent evolution of IL and SL studies

Ten years ago, it seemed possible to contrast IL and SL on their main issues of interest, namely syntax acquisition and lexicon formation, respectively. Indeed, the to-be-learned material used in artificial grammar learning research is typically governed by rules, that is by organizing principles which are independent of the specific material used in a given instance. If participants learned the rules, then this form of learning would be out of the scope of SL studies, in which the notion of rules is *a priori* irrelevant. However, research from the past few years has made it increasingly clear that participants in artificial grammar learning experiments do not need to extract the rules to perform well, even in situations involving transfer across surface forms (Box 1). In addition, the artificial grammar learning paradigms tend to be now supplanted by other paradigms, such as the serial reaction-time tasks, in which a description of the materials in terms of rules appears less appropriate.

Another initial difference between the two domains was that IL research used a large variety of situations involving different sensory modalities and response systems, whereas SL originally focused on the early stage of language acquisition. However, more recently research on SL has progressively broadened its scope of investigation. The syllables used in the first studies have been replaced by tones with the same results [11,12]. A parallel literature has evolved with visual shapes [6–8], or even tactile stimuli [13]. Perhaps even more importantly,

Corresponding author: Perruchet, P. (pierre.perruchet@u-bourgogne.fr).

Box 1. Fading out of the rule vs. no-rule

Transfer tasks, in which the form of the material presented during the training phase is changed, have been used in studies involving the learning of artificial materials by infants (e.g. [50]) and adults (e.g. [51]), and the learning of world-sized regularities by children and adults [52,53]. For instance, in the study by Marcus [50], infants previously exposed to exemplars of an ABB grammar (e.g. *ga-ti-ti*) subsequently listen more to the sentences generated by an ABA grammar (e.g. *wo-fe-wo*) than to sentences generated by the ABB grammar (e.g. *wo-fe-fe*), although new syllables were involved in both types of sentences.

Although the empirical evidence for the phenomenon is undisputed, the prevailing idea that positive transfer supports a rule-based interpretation has now been challenged. Some studies suggest that stimuli are not processed at an abstract level during incidental training [10]. The above chance performance of participants could be due to analogical processes triggered by the transfer items during the test [54], and could be mediated by the participants' explicit knowledge of the structure [55].

It is also possible to acknowledge some form of abstract coding without surmising rule-knowledge. Indeed, as argued by Redington and Chater [56], 'surface independence and rule-based knowledge are orthogonal concepts'. It is worth stressing that the evidence of transfer in implicit training conditions, even in adults, is limited to simple and salient features of the stimuli, and especially to the structure of repetitions (e.g. [57,58]). In these conditions, transfer might be based on the direct coding of abstract relations at the perceptual level [12,47].

recent SL studies are no longer limited to the segmentation of a continuous display into word-like units, but they also explore other, more complex structures [14]. For instance, Saffran and Wilson [15] have used a finite-state grammar to generate their artificial language, and Hunt and Aslin [16] used a serial reaction time task, hence borrowing the prototypical situations of IL to investigate the properties of SL.

A recent set of results on the role of attention further strengthens the similarity between IL and SL. Although a few earlier IL studies claimed that at least some forms of learning do not require attention, the bulk of recent evidence supports the opposite conclusion. For instance, Shanks and collaborators (e.g. [18]) showed that performances in serial reaction time tasks are degraded under double-task conditions (see also [19]). Likewise, Chun and Jiang (e.g. [20]) showed that implicit learning in the contextual cuing paradigm is robust only when relevant, predictive information is selectively attended to (see also [21]). In covariation learning, Hoffman and Seibald [22] showed that no learning occurs without attention, even when the to-be-learned covariations are highly salient (for reviews on earlier studies, see [23,24]). The same conclusion emerges from studies in SL. When the performance of participants in a dual task setting is compared with that of participants attending to the to-be-learned materials, the former is always degraded compared to the latter. This has been observed in standard word segmentation tasks [25] as well as in paradigms using visual shapes [8,26].

Thus arguably, IL and SL studies now pursue essentially the same objective - namely, the study of domain-general learning mechanisms acting on attended information in incidental, unsupervised learning situations (e.g. [17]).

A new question: chunk formation versus statistical computation

Although the similarities between IL and SL are impressive, comparing the interpretations favored in both fields leads us to a thought-provoking observation. In the IL literature, several models have been developed as alternatives to the initial rule-based view. The first alternative idea in artificial grammar learning research was that participants memorized the displayed strings of letters, then performed their grammaticality judgments on the basis of the similarity between the test items and the study items. The role of similarity in grammaticality judgments has been shown in some studies (e.g. [27]). However, there is also significant evidence that participants memorize fragments of strings, and that grammaticality judgments rely, at least partly, on this form of knowledge. It has been argued that the fragments or chunks provide a most efficient coding of the information, because learning makes their selection increasingly adapted to the structure of the material (Box 2). This kind of interpretation has been applied as well to other IL paradigms such as serial reaction-time tasks [28]

By contrast, the interpretation proposed in the SL approach postulates that participants perform statistical computations. Evidence for segmentation is generally attributed to the ability of participants to compute some kind of conditional probabilities between successive or contiguous elements. This interpretation prevails for auditory artificial languages (e.g. [14,15]) as well as for visual scenes (e.g. [7,9]). At the computational level, this interpretation is generally implemented by connectionist networks, most often SRNs (e.g. [29]).

Note that the contrast we draw here is not as clear-cut as our presentation suggests. There have been a few attempts to account for word segmentation with chunking models (e.g. [30]), although they have been virtually ignored in SL literature. More significantly, the performance in IL paradigms has been often simulated with SRNs (e.g. [31–33]). Because SRNs, like any connectionist network, are sensitive to statistical regularities, this means that certain IL researchers have construed implicit learning as statistical computations [3,32,34]. However, the coexistence of chunk-based theories and connectionist models within the IL literature has not drawn much attention, largely because their common opposition to rule-based models overshadowed their differences. The joint consideration of IL and SL studies now brings the contrast between the two accounts on the front of the scene.

Combining chunks and statistics: three possible scenarios

Nobody denies the existence of chunk knowledge. The advocates of statistical approaches claim themselves that learning shapes some kind of psychological units. For instance, Saffran and collaborators [15,35] have shown that training with unsegmented speech results in the formation of word-like units, rather than in strings of sounds the probability of which varies on a continuous dimension. Likewise, Baker and collaborators [26] and Fiser and Aslin [9] emphasize that the end result of SL

Box 2. Artificial grammar meets word segmentation

Figure 1 shows a finite-state grammar that has been widely used in IL studies. After exposure to strings of letters generated by this grammar (e.g. a or b below), participants are able to discriminate new grammatical and ungrammatical strings of letters.

It is now generally accepted that artificial grammar learning relies, at least partly, on the formation of small chunks. If those chunks were selected randomly, as illustrated in line (a) below, memorizing chunks would be difficult and inefficient: there would be many chunks, infrequently repeated across strings, and their recombination would have a low probability of generating a new grammatical string. Instead, learning consists in forming chunks that recur as often as possible in different strings, and whose recombination has high chance of generating a new grammatical string [59]. Line (b) below shows how the strings of letters displayed in line (a) can be segmented into more relevant units. For instance, the chunks now implicitly encode the recursive loop RFV, shown in red in the grammar diagram.

Note that the issue of artificial grammar learning framed in this way appears very close to that investigated in word segmentation studies. In both cases, learning consists of finding the most relevant units to encode information.

(a) SVD R F V S V R F V D X F V D R F V D X
 (b) S V D R F V S V R F V D X F V D R F V D X

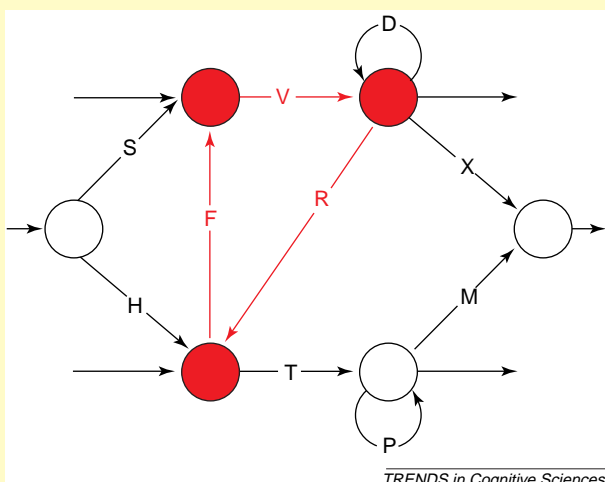


Figure 1. A typical finite-state grammar, with a recursive loop of letters, RFV, highlighted.

with visual displays is the formation of objects. This leaves three possibilities to account for the available evidence:

The first possibility is that statistical computations and chunk formation are independent processes. Meulemans and Van Der Linden [36] have argued for this view, with the additional assumption that chunk formation is responsible for conscious knowledge, and statistical computation for improved performance in implicit tasks (for related hypotheses, see [37,38]). This hypothesis is grounded on the dissociation between performance and explicit knowledge observed in amnesic patients. However, alternative interpretations have been proposed for this dissociation, notably by Shanks and collaborators [33,39,40]. Shanks and collaborators assume a single knowledge basis, and hence, in addition to the advantage of better parsimony, their interpretation provides a natural account for the ubiquitous relationships

observed in normal participants between conscious knowledge and performance.

The second possibility is that statistical computations and chunk formation are two successive steps in the learning process. Chunks would be inferred from prior statistical computations. Typically, chunk boundaries are defined as the points where the predictability of successive or spatially contiguous elements is the lowest. This interpretation is largely prevalent in SL research, both for oral stimuli (e.g. [35]) and for visual scenes (e.g. [9]).

The third possibility is that the formation of chunks is the only effective process, with the sensitivity to statistical structure being a by-product of this process. At least two computationally implemented models illustrate this option, the Competitive Chunking model [41] and *PARSER* [30]. In *PARSER*, for instance, the chunks are formed from the outset on a random basis, as a natural consequence of the capacity-limited attentional processing of the incoming information. These chunks are then forgotten or strengthened according to the laws governing associative memory.

We will now focus on the last two possibilities, the first in which chunking is based on prior statistical analyses, and the second in which chunking is a primitive process the result of which amounts to simulating statistical computations.

Does efficient chunking need prior statistical computations?

We are not aware of empirical arguments from proponents of chunk-based theories against the models assuming statistical computations, except that this assumption could be unnecessary. By contrast, SL researchers have occasionally argued that chunk models are only sensitive to the raw frequency of co-occurrences [14], whereas studies in SL have shown that participants were sensitive to more subtle statistics, such as conditional (or transitional) probabilities (e.g. [14,42]). Indeed, most chunk-based learning models, such as the competitive chunking model [41] or the measures of chunk strength used in the influential studies by Knowlton and collaborators (e.g. [43]) exclusively exploit frequency information, certainly because this initially appeared to be sufficient to account for a large part of the available evidence.

However, the exclusive focus on frequency of many chunk-based models is somewhat surprising in itself. Although chunk-based models are claimed to implement associative learning principles, assuming that chunk memory only depends on their frequency amounts to neglecting some of the most basic laws governing the formation of associations. Indeed, it has long been known that the strength of memory traces does not only depend on the number of repetitions of the study pairs. In particular, forgetting is due in large part to the interference generated by the prior or subsequent events that are related in some way to the target event. The sequential material used in both IL and SL studies is certainly prone to generate strong and pervasive interference, because it is typically generated by recombining a small number of primitives. Now, and this is the crucial point, taking into account the effect of interference in evaluating chunk

strength amounts to considering other measures of association than the raw frequency of co-occurrences. Box 3 illustrates how implementing forward interference is sufficient to make chunk strength sensitive to transitional probabilities, which SL researchers consider so important. Moreover, Perruchet and Peereman [44] have shown that *PARSER*, thanks to the role ascribed to interference in chunk formation, was even sensitive to contingency, that is to a measure of association more comprehensive than conditional probabilities.

The above remarks suggest that it might turn out to be difficult to decide between concurrent interpretations based on a simple consideration of their explanatory power. Because IL and SL have mainly evolved as separate fields of research, the challenge has not often been addressed. A few recent studies, however, have begun to explore situations designed in such a way that predictions drawn from chunk-based models and statistical approaches differ. In these studies, some version of an SRN is used to quantify the predictions of statistical approaches, whereas the chunking models are represented by the Competitive Chunking Model [45] or *PARSER* [37,44,46]. Although a detailed review of these studies is beyond the scope of this review, suffice it to say that, overall, their results do not clearly favor one or the other account.

These preliminary results suggest that the present accounts will need to be amended. Further models would also allow to encompass data that neither the chunk-based models nor the statistical approaches in their current instantiations seem to be able to explain. In the past few years, several studies have shown the possibility of incidentally learning the relations between elements that are not contiguous in space and/or time

(Box 4). Because the chunking process is usually construed as the clustering of adjacent events, these data confront the chunking models with a difficult challenge, as noted by Kuhn and Dienes [47]. In principle, they do not raise the same problem for statistical approaches, because the notion of statistical computations does not care about the nature of the data (e.g. contiguous or not) on which statistics may be computed. However, there is a consensus among researchers working on language and visual perception that models relying on statistical computations alone need to be constrained to avoid combinatorial explosion. The adjacency of the to-be-learned elements provides such a natural constraint (which is implemented, for instance, in SRNs). Thus, the possibility of learning non-adjacent dependencies entails either an in-depth revision of chunk-based models, or a significant departure from the most frequent computational implementation of statistical approaches.

Implications for the issue of consciousness

One of the major implication of the debate outlined above is the function of consciousness in the learning process. If the chunks are inferred from the results of statistical computations, then most of the learning process must be thought of as unconscious, because statistical computations are not performed consciously in the context of incidental learning paradigms. Of course, this does not mean that chunks, once formed, are functionally inert in further steps of conscious activities, but simply that their initial emergence is guided by unconscious computations. On the other hand, if the final chunks evolve from the progressive modification of primitive chunks, then the function of consciousness in chunk formation can be

Box 3. Statistical computations and chunk-based models: how do they converge towards the same predictions?

A B G H C D E F A B G H A B E F C D G H

Above is a 20-letter sequence made up from 8 different letters. Let us assume that they stand for syllables (although they could equally stand for tones of different pitches, the consonant letters typically used in artificial grammar learning studies, the locations of a target on a screen typically involved in serial reaction-time studies, or any other events). The sequence can be viewed as the random succession of four bisyllabic words (they have been colored for ease of reading). How can the words be discovered?

One solution consists of considering the frequency of all the bisyllabic units. However, column (a) of Table I shows that, because AB and GH are more frequent than the other words, the 'part-word' BG turns out to be as frequent as the 'words' CD and EF. Aslin and collaborators [42,6,7,16] used a similar design to show that participants do not exploit co-occurrence frequencies, but rather the Transitional Probabilities (TP: Prob. y/x = frequency of xy /frequency of x). Indeed, as indicated in column (b), considering TPs solves the problem (all word-internal TPs are stronger than TPs straddling word boundaries), hence the prevalent claim in the SL literature that participants compute TP.

However, as shown in column (c), the same result can emerge if one considers instead that participants memorize chunks, as in IL studies. If memory for chunks was dependent only on their frequency, values in (c) would be identical to values in (a). However, memory consolidation and forgetting also depends on interference. Classical studies on interference show that the memory for AB is impaired by the presentation of AC or AD. For the sake of illustration, we have assumed

here that each occurrence of AB strengthens it by 1 unit, and each occurrence of another letter pair beginning with A decreases the AB strength by 0.5 unit. These parameters were selected arbitrarily, but the crucial outcome – namely that all the words have a stronger strength than any part-word – remains true whatever the parameters (the Pearson r between (b) and (c) is 0.95).

Table I. Analysis of the letter sequence shown above

	(a)		(b)	(c)
Units	Frequency		TP	Chunk strength
	xy	x	xy/x	$xy - ((x - xy) * 0.5)$
AB	3	3	1	3
CD	2	2	1	2
EF	2	2	1	2
GH	3	3	1	3
BE	1	3	0.33	0
BG	2	3	0.67	1.5
DE	1	2	0.5	0.5
DG	1	2	0.5	0.5
FA	1	2	0.5	0.5
FC	1	2	0.5	0.5
HA	1	2	0.5	0.5
HC	1	2	0.5	0.5

Box 4. Learning non-adjacent dependencies

Both IL and SL approaches have focused on the human ability to detect and exploit the relations between elements in close temporal or spatial proximity. However, linguistic structures, as well as other domains of high-level knowledge such as music, also include remote dependencies. That is to say, a relation exists between two events, A and C, irrespective of the intervening events (X), as illustrated in (a) below. In the past few years, this has given rise to a set of studies investigating the possibility of learning non-adjacent dependencies in artificial languages [60–63] and in music [64,65] in incidental conditions.

The results show a consensus that learning non-adjacent dependencies is possible, but under far more restrictive conditions than those required for learning the relations between contiguous events. Gomez [60,66] showed that the degree to which the AXC relationships are learned depends on the variability of the middle element (X). For Newport and Aslin (e.g. [61]) and Onnis *et al.* [62], the crucial factor is the similarity between A and C. Another facilitating condition is that the AXC units are displayed as individualized, pre-segmented units, rather than embedded within a continuous stream of stimuli.

Closely related are studies in which the to-be learned stimuli are generated by a bi-conditional grammar [40,47], as in (b), or in which the dependencies are self-embedded [67,68,4], as in (c). In these more complex cases, the possibility of obtaining evidence of learning in truly incidental conditions does not seem clearly established yet.

(a) 

Non adjacent dependency: The relation is between A and C, whatever X

(b) 

Biconditional grammar: The relations are between A and X, B and Y, etc.

(c) 

Self embedded dependencies: The relations are between A and X, B and Y, etc.

construed differently. Starting from the postulate that chunks are the actual content of phenomenal experience, Perruchet and Vinter [48,49] outlined a view of the human mind in which consciousness is thought of as self-organized. In this model, the optimal coding of the incoming information occurs as a natural by-product of the evolution of conscious percepts and representations, under the action of simple associative learning and memory processes.

Conclusion

Recent evolution of research on both IL, initially aimed at studying rule abstraction in complex situations, and SL, initially focused on word segmentation, suggests that the two lines of research explore the same domain-general incidental learning processes. Bringing together these two domains of research, however, reveals a divergence between the interpretation favored in IL, which focuses on the formation of chunks, and the interpretation favored in SL, which relies on statistical computations. One possibility is that chunks are inferred from the results of (unconscious) statistical computations. Another possibility is that (perhaps conscious) chunks are formed from the outset and then evolve as a result of basic associative learning principles. It is clear that there are many challenges for future research in these two areas (see Box 5).

Box 5. Questions for future research

- Bringing together IL and SL domains of research suggests many directions for further studies, because each domain has explored a limited set of issues, and it appears necessary to confirm that the conclusions reached in one domain may be generalized to the experimental paradigms typical of the other one. The following two questions instantiate this research strategy.
- A large part of the recent IL literature is concerned with the relative invariance of implicit learning mechanisms with age, in children and elderly people. Furthermore, several studies have investigated whether these mechanisms are preserved in people with mental retardation, and in patients with psychiatric and neurological disorders. Are results the same with the paradigms generally used in SL studies? Likewise, there has been some brain imaging studies of IL, and investigating whether the same neural regions would mediate SL would be of interest.
- Although both IL and SL studies involve arbitrary materials, it turns out that these materials can never be considered to be completely neutral with regards to learning. For instance, participants may be guided towards (or diverted from) the discovery of its underlying structure by prior knowledge of related real situations, the acoustic or visual features of the materials, and so on. In SL research, the interactions between the statistical structure and the other features of the situations that might affect learning have been taken as a valuable problem of its own, especially in the context of lexicon formation in young children. Similar problems should be explored with the paradigms used in IL research.
- Are conditional probabilities the main statistic to which human behavior is sensitive, as the literature on SL assumes? Research on conditioning has suggested that conditioned performances depend on more elaborate measures of association, such as Delta P or contingency. IL/SL research on this question has hardly scratched the surface in this area.
- Irrespective of the importance given to chunks in the dynamics of learning, it seems essential to investigate this notion further. For instance, can chunks be thought of as the content of the attentional focus, or is it possible to assume the existence of functional chunks that participants would not be aware of?

Acknowledgements

This work was supported by grants from the Centre National de la Recherche Scientifique (CNRS, UMR 5022 and FRE 2987), from the Université de Bourgogne, from the Région de Bourgogne (AAFE), and from the Université Paris V. The authors thank Stephanie Chambaron, Suzanne Filipic, Bob French, Barbara Tillmann, and the anonymous reviewers of a first draft for their help at various stages of elaboration.

References

- 1 Reber, A.S. (1967) Implicit learning of artificial grammars. *J. Verbal Learn. Verbal Behav.* 6, 855–863
- 2 Reber, A.S. (1993) *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*, Oxford University Press
- 3 Cleeremans, A. *et al.* (1998) Implicit learning: News from the front. *Trends Cogn. Sci.* 2, 406–416
- 4 Shanks, D.R. (2005) Implicit learning. In *Handbook of Cognition* (Lamberts, K. and Goldstone, R., eds), pp. 202–220, Sage Publications
- 5 Saffran, J.R. *et al.* (1996) Statistical learning by 8-month-old infants. *Science* 274, 1926–1928
- 6 Fiser, J. and Aslin, R.N. (2001) Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol. Sci.* 12, 499–504
- 7 Fiser, J. and Aslin, R.N. (2002) Statistical learning of higher-order temporal structure from visual shape sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 458–467
- 8 Turk-Browne, N.B. *et al.* (2005) The automaticity of visual statistical learning. *J. Exp. Psychol. Gen.* 134, 552–564
- 9 Fiser, J. and Aslin, R.N. (2005) Encoding multielement scenes: Statistical learning of visual features hierarchies. *J. Exp. Psychol. Gen.* 134, 521–537

- 10 Conway, C.M. and Christiansen, M.H. Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychol. Sci.* (in press)
- 11 Saffran, J.R. *et al.* (1999) Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52
- 12 Saffran, J.R. *et al.* (2005) Changing the tune: Absolute and relative pitch processing by adults and infants. *Dev. Sci.* 8, 1–7
- 13 Conway, C.M. and Christiansen, M.H. (2005) Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 24–39
- 14 Saffran, J.R. (2001) The use of predictive dependencies in language learning. *J. Mem. Lang.* 44, 493–515
- 15 Saffran, J.R. and Wilson, D.P. (2003) From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy* 4, 273–284
- 16 Hunt, R.H. and Aslin, R.N. (2001) Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *J. Exp. Psychol. Gen.* 130, 658–680
- 17 Kirkham, N.Z. *et al.* (2002) Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition* 83, B35–B42
- 18 Shanks, D.R. *et al.* (2005) Attentional load and implicit sequence learning. *Psychol. Res.* 69, 369–382
- 19 Remillard, G. (2003) Pure perceptual-based sequence learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 581–597
- 20 Jiang, Y. and Chun, M.M. (2001) Selective attention modulates implicit learning. *Q. J. Exp. Psychol.* 54A, 1105–1124
- 21 Jiang, Y. and Leung, A.-W. (2005) Implicit learning of ignored visual context. *Psychon. Bull. Rev.* 12, 100–106
- 22 Hoffmann, J. and Sebald, A. (2005) When obvious covariations are not even learned implicitly. *Eur. J. Cog. Psychol* 17, 449–480
- 23 Shanks, D.R. (2003) Attention and awareness in ‘implicit’ sequence learning. In *Attention and Implicit Learning* (Jiménez, L., ed.), pp. 11–42, John Benjamins
- 24 Hsiao, A.T. and Reber, A.S. (1998) The role of attention on implicit sequence learning. In *Handbook of Implicit Learning* (Stadler, M.A. and Frensch, P., eds), pp. 471–494, Sage Publications
- 25 Toro, J.M. *et al.* (2005) Speech segmentation by statistical learning depends on attention. *Cognition* 97, B25–B34
- 26 Baker, C.I. *et al.* (2004) Role of attention and perceptual grouping in visual statistical learning. *Psychol. Sci.* 15, 460–466
- 27 Pothos, E.M. and Bailey, T.M. (2000) The role of similarity in artificial grammar learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 847–862
- 28 Buchner, A. *et al.* (1998) On the role of fragmentary knowledge in a sequence learning task. *Q. J. Exp. Psychol.* 51A, 251–281
- 29 Christiansen, M.H. *et al.* (1998) Learning to segment speech using multiple cues: A connectionist model. *Lang. Cogn. Processes* 13, 221–268
- 30 Perruchet, P. and Vinter, A. (1998) PARSER: A model for word segmentation. *J. Mem. Lang.* 39, 246–263
- 31 Cleeremans, A. and McClelland, J.L. (1991) Learning the structure of event sequences. *J. Exp. Psychol. Gen.* 120, 235–253
- 32 Cleeremans, A. (1993) *Mechanims of Implicit Learning: A Connectionist Model of Sequence Processing*, MIT Press
- 33 Kinder, A. and Shanks, D.R. (2003) Neuropsychological dissociations between priming and recognition: a single-system connectionist account. *Psychol. Rev.* 110, 728–744
- 34 Tillmann, B. *et al.* (2000) Implicit learning of tonality: A self-organizing approach. *Psychol. Rev.* 107, 885–913
- 35 Saffran, J.R. (2001) Words in a sea of sounds: The output of statistical learning. *Cognition* 81, 149–169
- 36 Meulemans, T. and Van der Linden, M. (2003) Implicit learning of complex information in amnesia. *Brain Cogn.* 52, 250–257
- 37 Jimenez, L. (2005) Chunk structure in implicit and explicit sequence learning, *2nd European Workshop on Movement Science* (Vienna), Abstract No. 2.1.6
- 38 Anderson, J.R. and Lebiere, C. (1998) *The Atomic Components of Thought*, Erlbaum
- 39 Shanks, D.R. *et al.* Disruption of sequential priming in organic and pharmacological amnesia: A role for the medial temporal lobes in implicit contextual learning. *Neuropsychopharmacology* (in press)
- 40 Shanks, D.R. *et al.* (2002) Modularity and artificial grammar learning. In *Implicit Learning and Consciousness* (French, R. and Cleeremans, A., eds), pp. 93–120, Psychology Press
- 41 Servan-Schreiber, D. and Anderson, J.R. (1990) Learning artificial grammars with competitive chunking. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 592–608
- 42 Aslin, R.N. *et al.* (1998) Computation of conditional probability statistics by 8-month-old infants. *Psychol. Sci.* 9, 321–324
- 43 Chang, G.Y. and Knowlton, B.J. (2004) Visual feature learning in artificial grammar classification. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 714–722
- 44 Perruchet, P. and Peereman, R. (2004) The exploitation of distributional information in syllable processing. *J. Neuroling.* 17, 97–119
- 45 Boucher, L. and Dienes, Z. (2003) Two ways of learning associations. *Cogn. Sci.* 27, 807–842
- 46 Giroux, I. and Rey, A. (2005) Word and sub-word units in speech perception. *Proceedings of the 46th Annual Meeting of the Psychonomic Society* (Toronto), Abstract No. 3061
- 47 Kuhn, G. and Dienes, Z. Implicit learning of non-local musical rules. *J. Exp. Psychol. Learn. Mem. Cogn.* (in press)
- 48 Perruchet, P. (2005) Statistical approaches to language acquisition and the self-organizing consciousness: A reversal of perspective. *Psychol. Res.* 69, 316–329
- 49 Perruchet, P. and Vinter, A. (2002) The self-organizing consciousness. *Behav. Brain Sci.* 25, 297–388
- 50 Marcus, G.F. *et al.* (1999) Rule learning by seven-month-old infants. *Science* 283, 77–80
- 51 Dienes, Z. and Altmann, G. (1997) Transfer of implicit knowledge across domains: How implicit and how abstract?. In *How Implicit is Implicit Learning?* (Berry, D., ed.), pp. 107–123, Oxford University Press
- 52 Pacton, S. *et al.* (2001) Implicit learning out of the lab: The case of orthographic regularities. *J. Exp. Psychol. Gen.* 130, 401–426
- 53 Pacton, S. *et al.* (2005) Children’s implicit learning of graphotactic and morphological regularities. *Child Dev.* 76, 324–339
- 54 Vokey, J.R. and Higham, P.A. (2005) Abstract analogies and positive transfer in artificial grammar learning. *Can. J. Exp. Psychol.* 59, 54–61
- 55 Gomez, R.L. (1997) Transfer and complexity in artificial grammar learning. *Cogn. Psychol.* 33, 154–207
- 56 Redington, M. and Chater, N. (2002) Knowledge representation and transfer in artificial grammar learning. In *Implicit Learning and Consciousness* (French, R. and Cleeremans, A., eds), pp. 121–143, Psychology Press
- 57 Gomez, R.L. *et al.* (2000) The basis of transfer in artificial grammar learning. *Mem. Cogn.* 28, 253–263
- 58 Tunney, R.J. and Altmann, G.T.M. (2001) Two modes of transfer in artificial grammar learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 614–639
- 59 Perruchet, P. *et al.* (2002) The formation of structurally relevant units in artificial grammar learning. *Q. J. Exp. Psychol.* 55A, 485–503
- 60 Gomez, R. (2002) Variability and detection of invariant structure. *Psychol. Sci.* 13, 431–436
- 61 Newport, E.L. and Aslin, R.N. (2004) Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cogn. Psychol.* 48, 127–162
- 62 Onnis, L. *et al.* (2005) Phonology impacts segmentation in online speech processing. *J. Mem. Lang.* 53, 225–237
- 63 Perruchet, P. *et al.* (2004) Learning non-adjacent dependencies: No need for algebraic-like computations. *J. Exp. Psychol. Gen.* 133, 573–583
- 64 Creel, S.C. *et al.* (2004) Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 1119–1130
- 65 Dienes, Z. and Longuet-Higgins, C. (2004) Can musical transformations be implicitly learned? *Cogn. Sci.* 28, 531–558
- 66 Gomez, R.L. and Maye, J. (2005) The developmental trajectory of nonadjacent dependency learning. *Infancy* 7, 183–206
- 67 Poletiek, F.H. (2002) Implicit learning of a recursive rule in an artificial grammar. *Acta Psychol. (Amst.)* 111, 323–335
- 68 Perruchet, P. and Rey, A. (2005) Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bull. Rev.* 12, 307–313