Consciousness and Cognition 28 (2014) 1–16

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Role of prior knowledge in implicit and explicit learning of artificial grammars ☆

Eleni Ziori [a,*], Emmanuel M. Pothos [b], Zoltán Dienes [c]

[a] Department of Psychology, University of Ioannina, Greece
[b] Department of Psychology, City University London, UK
[c] Sackler Centre for Consciousness Science and School of Psychology, University of Sussex, Brighton, UK

ARTICLE INFO

Article history:
Received 27 July 2013

Keywords:
AGL
Implicit learning
Explicit learning
General knowledge
Similarity
Rules

ABSTRACT

Artificial grammar learning (AGL) performance reflects both implicit and explicit processes and has typically been modeled without incorporating any influence from general world knowledge. Our research provides a systematic investigation of the implicit vs. explicit nature of general knowledge and its interaction with knowledge types investigated by past AGL research (i.e., rule- and similarity-based knowledge). In an AGL experiment, a general knowledge manipulation involved expectations being either congruent or incongruent with training stimulus structure. Inconsistent observations paradoxically led to an advantage in structural knowledge and in the use of general world knowledge in both explicit (conscious) and implicit (unconscious) cases (as assessed by subjective measures). The above findings were obtained under conditions of reduced processing time and impaired executive resources. Key findings from our work are that implicit AGL can clearly be affected by general knowledge, and implicit learning can be enhanced by the violation of expectations.

1. Introduction

The present work uses the AGL paradigm to explore the effect of prior knowledge on (implicit and explicit) learning. Most AGL studies use meaningless stimuli, devoid of any correspondence with prior knowledge. Moreover, the majority of AGL models reference only the structural aspects of stimuli (e.g., Boucher & Dienes, 2003; Cleeremans, 1993a, 1993b; Dienes, Altmann, & Gao, 1999; Servan-Schreiber & Anderson, 1990; but see Dienes & Fahey, 1995; Sun, 2000). Extending the AGL paradigm to a knowledge-rich version is crucial in determining whether AGL theory can extend to more realistic learning conditions and whether AGL tasks can be employed to shed light on how general knowledge can influence cognitive processes.

In a typical AGL experiment (e.g., Dulany, Carlson, & Dewey, 1984; Reber, 1967; Reber & Allen, 1978), participants first study a list of letter strings generated by a finite state grammar and are asked to simply observe them or memorize them.

☆ Emmanuel M. Pothos was supported by Leverhulme Trust grant RPG-2013-004 and by Air Force Office of Scientific Research (AFOSR), Air Force Material Command, USAF, grant FA 8655-13-1-3044. The U.S Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright notation thereon.

* Corresponding author. Address: Department of Psychology, Faculty of Philosophy, Education & Psychology, School of Philosophy, University of Ioannina, Dourouti, 451 10 Ioannina, Greece. Fax: +30 26510 05805.

E-mail address: eziori@cc.uoi.gr (E. Ziori).

http://dx.doi.org/10.1016/j.concog.2014.06.003
1053-8100/

After training, they are informed that the strings followed a complex set of rules, but no specific information is provided regarding the nature of the rules. Then, they are asked to classify new letter strings, half of which are consistent with the rules and are thus called grammatical (G) and half of which are not consistent with the rules and are called non-grammatical (NG). No corrective feedback is provided in the test phase.

Replacing the letter strings in the standard AGL paradigm with meaningful stimuli (e.g., sequences of cities), without any more elaborate prior knowledge manipulation, does not seem to alter performance (Pothos, Chater, & Ziori, 2006). Pothos (2005, Experiment 2) used sequences of cities and also manipulated the consistency of stimulus structure with instructions given to participants, to induce different expectations about the stimuli. In his relevant experiment, the stimuli were sequences of cities that corresponded to the routes of a salesman. In one condition, stimulus structure was consistent with participants' expectations from the instructions (that the salesman should make as many short trips as possible), whereas in the other it was not. When stimulus structure was inconsistent with expectations, performance was impaired. At the very least, the study of Pothos (2005) shows that expectations about stimulus structure can affect AGL performance. However, Pothos (2005) used only a simple incidental learning condition and measured AGL performance only in terms of grammaticality accuracy. Further, Pothos did not employ any measures of the implicitness of the acquired knowledge. The present work extends Pothos's (2005) study, to explore the generality of his findings. In particular, we disentangle three knowledge types, namely general knowledge relations that are consistent or inconsistent with people's expectations and two purely structural aspects (i.e., grammaticality and similarity). We also use subjective measures of implicitness, to examine the implicit or explicit nature of each knowledge type, under different learning conditions.

1.1. Key features of AGL

A key aspect of AGL tasks concerns stimulus construction. Consider, for example, Fig. 1, which presents the grammar employed in the present experiment.

In going from left to right, a set of G strings are created, as opposed to NG strings. The distinction between G and NG sequences is referred to as 'grammaticality'. Crucially, the relation between the training and test stimuli is not limited to the grammar rules per se: For example, some test items will have bigrams (pairs of symbols) or trigrams (triplets of symbols), which are familiar from training (Perruchet & Pacteau, 1990; Knowlton & Squire, 1996).

There has been a flourishing research tradition and divergent hypotheses on what is learned in AGL, including rules, whole item similarity and similarity based on chunk overlap (e.g., Dulany et al., 1984; Knowlton & Squire, 1996; Perruchet & Pacteau 1990; Reber, 1967; Reber, 1989; Vokey & Brooks, 1992; see Pothos, 2007 for a review). For example, apart from grammaticality, a common measure that has been used in AGL is Knowlton and Squire's (1996) chunk strength index. Chunk strength is estimated by averaging the frequency, with which all chunks (i.e., bigrams or trigrams) of each test item occurred during training (cf. Meulemans & Van der Linder, 1997). When grammaticality and chunk strength are carefully balanced, as is the case in the present work, the former can be thought to constitute more rule-like knowledge (in the sense that it does not depend on frequency, at least of chunks) and the latter more similarity-like knowledge. Henceforth, when we say grammaticality we will imply rule-like knowledge over and above the frequency-dependent distributional characteristics of chunks, and when we refer to similarity we will imply Knowlton and Squire's (1996) chunk-strength index.

AGL has also been widely used in the implicit learning literature. The issue of the implicitness of knowledge has been hotly debated (e.g., Dulany, 2003; Perruchet & Pacteau, 1990; Shanks & St. John, 1994). In this work, we adopt a general definition, according to which *implicit learning* refers to learning without the need for intention (i.e., it can be passive), and

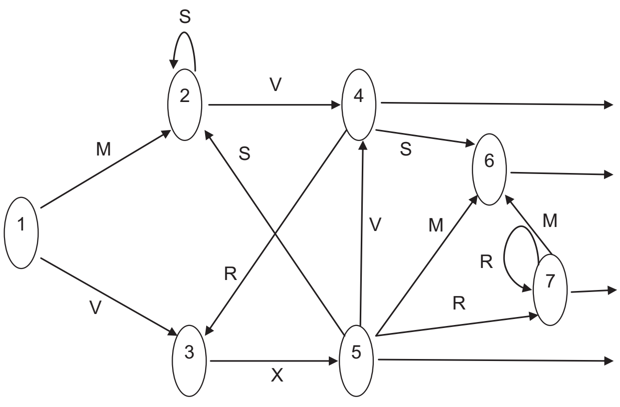

Fig. 1. The present grammar is a deterministic version of Reber and Allen's (1978) classic grammar, though the selection for training and test items was based on the procedure of Bailey and Pothos (2008). *Note*: A nondeterministic grammar can have two (or more) arrows out of a given state that have the same label on each arrow. A deterministic grammar has different labels on all the arrows out of any given state, so if you know what state you are in and what the next symbol is, you know with certainty what the next state is.

results in knowledge that has some content not directly available to conscious introspection (for a review of conscious vs. unconscious processes see Norman, 2010).

During testing in AGL tasks, participants may rely on any type of structural knowledge (e.g., bigrams, rules, whole items) they acquired during training. In this study, the implicitness of the *structural* knowledge was assessed with Dienes and Scott's (2005) method, which asks participants to specify the basis of their classifications, by choosing among different knowledge attributions (e.g., intuition, memory; Guo et al., 2011; Hamrick & Rebuschat, 2012; Neil & Higham, 2012; Norman & Price, 2012). These attributions allowed us to assess whether, when each of our three knowledge types (i.e., general knowledge relations, grammaticality, chunk strength) influenced participants' judgments, there was explicit (conscious) knowledge of what that structure was or that it was a particular structure motivating the judgment.

AGL research has led to conflicting conclusions regarding the implicitness/explicitness of similarity and grammaticality, with some researchers (Chang & Knowlton, 2004; Higham, 1997) associating similarity-based learning with an intentional, conscious and effortful process and the learning of grammaticality with a more automatic and passive process, and others (Pothos & Wood, 2009) suggesting the opposite. With respect to the implicitness of general knowledge relations, different views have also been proposed. On one view, theory-based knowledge is considered explicit (e.g., McRae, 2004; Pothos, 2005; Sloman, 1996). On this view, we would expect general knowledge relations in the present study to be associated with explicit knowledge. On other views, the influence of prior knowledge may be considered as the influence of prior exemplars (Heit, 1994), and, thus, might include some implicit memories. Most importantly, the implicitness results would bear on the important debate of whether prior knowledge is an unselective and passive process with no room for possible interpretative biases (Hayes & Broadbent, 1988) or it is a selective process (van den Bos & Poletiek, 2009) that can interact with expectations (e.g., Sun, 2000; Sun, Merrill, & Peterson, 2001; Ziori & Dienes, 2008).

In sum, the present work aims to explore AGL performance, by disentangling three different knowledge types (i.e., grammaticality, similarity and one that reflected general knowledge) and examining their implicit or explicit nature. We so hope to provide a clearer picture of the effect of prior knowledge on AGL performance.

1.2. The effect of prior knowledge on learning

The effect of prior knowledge on learning has been ignored in AGL research. By contrast, in categorization research, the possibly profound influence of general knowledge on performance has been widely appreciated (Harris, Murphy, & Rehder, 2008; Kaplan & Murphy, 2000; Medin, Wattenmaker, & Hampson, 1987; Murphy & Allopenna, 1994; Murphy & Kaplan, 2000; Murphy & Wisniewski, 1989; Pazzani, 1991; Wattenmaker, Dewey, Murphy, & Medin, 1986; Wisniewski, 1995).

A clear conclusion from the bulk of categorization research is that congruency of prior knowledge with stimulus structure, typically, facilitates learning. What is less clear is what happens when prior knowledge is incongruent with the stimuli. According to conflict monitoring theory (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001; Verguts & Notebaert, 2009), when a stimulus or a situation involves conflict, the cognitive system allocates more attentional resources to adapt to the relevant situation, which, in turn, reduces the influence of irrelevant information. According to the predictive coding approach, stimuli are coded precisely in terms of the extent to which they violate expectations (Clark, 2013). Similarly, in categorization research, Heit (1998) found that, under slow-paced learning, observations incongruent with prior knowledge affected categorization more than congruent observations did. Conversely, Pothos (2005) found that incongruence of prior knowledge with stimulus structure impaired AGL grammaticality performance.

We aim to test the generality of Pothos's (2005) conclusion, by examining how prior knowledge affects grammaticality together with other knowledge types, and under different learning conditions. The guiding question is this: Does incongruence of prior knowledge with structure interfere with performance of all or some knowledge types and under all or specific learning conditions?

1.3. Logic of the current investigation

As in Pothos (2005), the current AGL task employed sequences of city names as stimuli, but we also extend this work in several ways. Training stimuli were constructed to embody an association between high-frequency bigrams and shorter inter-city distances and were the same for all participants. A knowledge type that characterized stimuli, the so-called 'distance' factor, was computed from an approximate consideration of whether the city transitions in each test stimulus corresponded to shorter or longer distances. Performance on this factor reflected participants' prior (geographical) knowledge. A key experimental condition concerned instructions that elicited different expectations about the stimuli. Participants were told that the stimuli were routes of an airline company and it was stated that it was advantageous for the company to either make more trips to nearby cities (the consistent condition) or to far cities (the inconsistent condition). In the consistent condition, distance accuracy reflected the degree to which participants endorsed more short-distance items, in line with both their expectations (derived from the cover story) and the training stimulus structure (which reflected shorter inter-city distances). Conversely, in the inconsistent condition, distance accuracy reflected the degree to which participants endorsed more long-distance items over short-distance ones, in line with the expectations induced from the cover story, but at odds with the training stimulus structure. By comparing performance in these two conditions, we can examine how the congruency or incongruency of prior knowledge can affect learning (cf. Pothos, 2005, who provided the same instructions to all participants, but manipulated stimulus structure, to be consistent or inconsistent with expectations).

Our test stimuli were balanced along three knowledge factors, the distance factor, grammaticality, and similarity (as chunk strength). This design illustrates how our AGL task extends Pothos's (2005) work and goes beyond similar work in categorization, as, in the present work, performance during testing can be influenced by three theoretically-relevant performance factors, which are pairwise (nearly) orthogonal to each other. By contrast, Pothos (2005) assessed performance only in terms of grammaticality (and only in a baseline learning condition).

Another innovation of our work concerns the inclusion of two learning conditions, beyond the baseline one. The first condition was time pressure, whereby exposure to the training stimuli was greatly reduced. This condition allows us to distinguish between performance factors that require *time* to develop and ones that do not. For example, perceptual categorization research has shown that the application of background knowledge can be fast (Lin & Murphy, 1997; Luhmann, Ahn, & Palmeri, 2006; Palmeri & Blalock, 2000). Other researchers (Sloman, 1996; Smith & Sloman, 1994) equate knowledge-based categorization with rule-based processing, which they consider as time and effort demanding and contrast it with similarity-based categorization, which they consider automatic. Cleeremans and Jiménez (2002) suggested that representations become explicit depending on their quality and that one factor affecting quality is stability in time. Thus, on this view, under time pressure, we expect performance to involve mostly implicit knowledge (cf. Mealor & Dienes, 2012; but see Mealor & Dienes, 2013).

The second learning condition was a concurrent task (a working memory load) in training. A straightforward assumption is that such tasks engage executive control (Baddeley, 2007; Baddeley & Andrade, 1998; Shelton, Elliott, & Cowan, 2008). Accordingly, if a performance factor is not affected by the concurrent task, then we can infer that the supporting process does not require executive resources and is so more automatic (Kemler Nelson, 1984; Smith & Shapiro, 1989). Moreover, various investigators have shown that a concurrent task is more likely to interfere with explicit learning, than with implicit learning (e.g., Dienes, Altmann, Kwan, & Goode, 1995; Dienes & Scott, 2005; Roberts & MacLeod, 1995; Waldron & Ashby, 2001; Ziori & Dienes, 2008; contrast Shanks & Channon, 2002).

In sum, our work derives from Pothos's (2005) study, but provides several key extensions allowing the study of different knowledge factors, the implicitness/explicitness of the acquired knowledge, and the dependence of learning on time pressure and a concurrent task. We thus hope to provide a more comprehensive examination of the complex issue of how prior knowledge affects learning.

2. Method

2.1. Participants/design

The experiment involved two between-participants conditions, consistency (consistent vs. inconsistent) and training (baseline, time pressure and dual/concurrent task). One hundred and eighty participants, mostly undergraduate students at the University of Ioannina, Greece, were randomly assigned to one of the six training cells. The training stimuli were sequences of city names that corresponded to the routes of an airline company. Participants in both the consistent and the inconsistent conditions were presented with the same stimuli. However, instructions (and related general knowledge) were congruent with training stimulus structure in the consistent condition, and incongruent in the inconsistent condition.

2.2. Materials

Stimulus generation was based on a deterministic version of Reber and Allen's (1978) classic grammar. Stimuli had a length between three and seven city names. We used Bailey and Pothos' (2008) algorithm for generating AGL stimuli. There were 20 training items, and 40 test items, which were new (i.e., they had not been seen during training). Half of the test items were G and the other half NG. Chunk strength, which was based on Knowlton and Squire's (1996) index, was well counterbalanced in our stimulus set: The mean chunk strength of G and NG items were 4.25 (SD = 1.46) and 4.20 (SD = 1.13) respectively, resulting in a non-significant difference, p = .90, $\eta_p^2 < .001$.[1] Test items were ordered according to their chunk strength; there were 20 high-chunk strength items and 20 low-chunk strength items. The average chunk strength of the low-chunk strength items was 3.12 (SD = 0.51) and of the high-chunk strength items 5.33 (SD = 0.74), a significant difference, $t(38) = 10.94$, $p < .001$, $\eta_p^2 = .76$. Chunk strength values conformed roughly to a bimodal distribution, hence making the classification of test items into high/low chunk strength ones meaningful (Fig. 2).

Apart from the two purely structural factors of grammaticality and chunk strength, we also created a general knowledge factor, the 'distance factor'. We sought to identify a mapping between the symbols of the grammar and city names, such that more frequent training bigrams would correspond to cities closer together. It is in this way that the training items can be said to 'make sense' or not relative to the general (geographical) knowledge of participants and the added instructions that more frequent trips between nearby cities (in the consistent condition) or between far away cities (in the inconsistent condition) were advantageous.

[1] P values are reported here and elsewhere when characterizing the properties of the items, according to convention, but as the items used constituted the population, there is no statistical generalization to be made.

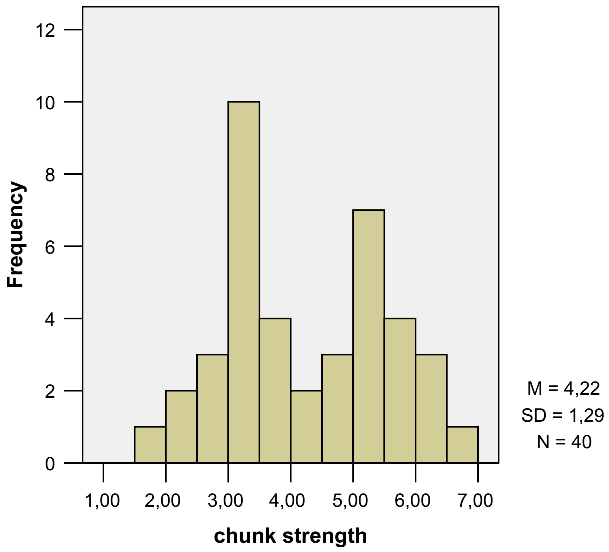

Fig. 2. The distribution of chunk strength values in the test items.

To create the distance factor, we first created a similarity matrix for all the symbols of the grammar (for brevity, we represent these as in the original Reber & Allen, 1978, study, as M, S, R, V, and X). To do so, we computed the frequency of bigrams in training items (Table 1) and the matrix entries corresponded to this information. For example, the frequency of bigram MS in training would be 10, therefore, this was the value in the corresponding entry in the similarity matrix. Bigram frequencies ranged between one and fourteen. We filled in the matrix elements as much as possible using distributional information from training, but five of the possible bigrams did not appear in training.

The value of each cell was subtracted from 14 (i.e., the maximum bigram frequency), so as to create a distance matrix from the similarity matrix, such that high bigram frequencies would be associated with short distances. Further, we assigned the value 0 to all the bigrams made of the same symbol (MM, SS etc.), since such bigrams should correspond to least distance. The missing cells were completed in either of two ways: First, if there was a distance for cell AB, but not BA, we set the value for BA to be the same as the value for AB (in cases where there was a value for AB different to BA, we replaced both values with their average). Second, where there was neither an AB value nor a BA one, we set both values to 7 (the median distance). In this way, we managed to create a distance matrix out of bigram frequencies, such that, *in general*, more frequent bigrams corresponded to shorter inter-city distances.

In order to provide a two-dimensional representation of the pattern of distances, we carried out a multidimensional scaling (MDS) analysis (Shepard, 1980). The MDS of the above distance matrix revealed a satisfactory (stress = .003, RSQ = .999) two-dimensional representation of the letters in the grammar (Fig. 3). Of course, the two dimensions cannot be interpreted in terms of any aspect of the grammar.

Using this MDS representation of the letters (Fig. 3), we assigned a distance 'value' to each bigram, to examine the relation of the distance factor to grammaticality and chunk strength. Table 1 shows these results and allows an appreciation of how

Table 1
The association between bigram frequency and whether it corresponds to a short, intermediate, or long distance in the MDS map.

Training bigrams	Frequency	Distance on the basis of the MDS map	Problems
VS	2	Intermediate	
RX	13	Intermediate	*
XR	6	Intermediate	
SV	14	Intermediate	*
MV	6	Long	
VX	4	Long	
RM	1	Long	
XM	1	Long	
XV	4	Long	
XS	4	Long	
MS	10	Short	
VR	13	Short	
RR	5	Short	*
SS	14	Short	

Note: A '*' indicates bigrams such that even though their frequency is high they correspond to a long distance (or vice versa).

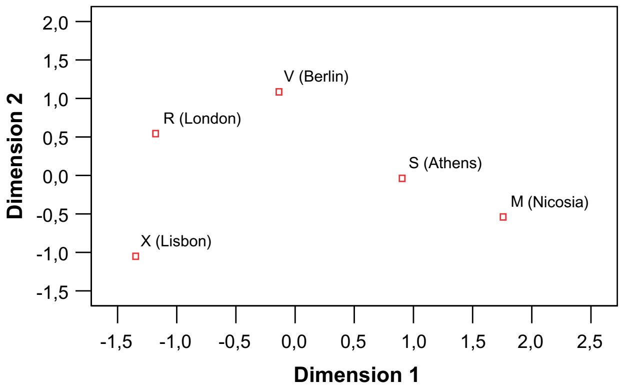

Fig. 3. An MDS representation of the letters in the Reber and Allen (1978) grammar, such that more frequent bigrams correspond to shorter distances.

successful the conversion of bigram frequencies to MDS distances has been: Overall, most of the frequent bigrams are indeed associated with shorter distances, but with some exceptions.

Given Table 1 association between distances and bigrams, we computed the distance factor for each test item. We assigned distance values of 0, 1, or 2 to each bigram, depending on whether the bigram corresponded to a short, intermediate, or long distance. Then, we computed the overall distance associated with a test item, by simply adding the distances corresponding to the bigrams of the test items. For example, consider test item MSVS. Its 'distance' can be found by adding the distance values for MS, SV, and VS, which is 0 + 1 + 1 = 2. The distance factor for the G and NG test items is shown in Table A1. Given the approximations involved in computing these distance values, we simply ordered the test items on the basis of distance and called 23 of them 'high distance' and 17 'low distance' (it was not possible to exactly equate the number of low and high distance items). Note that a dichotomous use of the distance variable takes into account our expectation that, at best, our participants' knowledge of geography would be approximate.

A key issue is whether the distance factor correlates with grammaticality or chunk strength. In both cases the answer is negative (for these correlations we used the continuous distance and chunk strength factors, not the dichotomized ones). The correlation with grammaticality was $r(40) = -.013$, and with chunk strength was $r(40) = .058$. Thus, the distance factor is largely independent of grammaticality and chunk strength. This conclusion is reinforced by comparing the distance factor for G and NG test items (4.85 vs. 4.90 respectively) and high and low chunk strength items (4.90 vs. 4.85 respectively). In sum, if participants were responding, predominantly, on the basis of distance, we would expect very low grammaticality and chunk strength performance.

The final issue concerned how to map letters to city names. First, we needed an arrangement of cities, which would conform as closely as possible to the arrangement of letters in Fig. 3. Second, the cities had to be such that their geographical locations would be broadly familiar to participants. Thus, we mapped the letters to the following European cities, that were expected to be familiar to our Greek participants: *M* was mapped to Nicosia, *S* to Athens, *V* to Berlin, *R* to London, and *X* to Lisbon. This mapping between letters and city names captures the two-dimensional MDS representation of the letters in the grammar satisfactorily (Fig. 4).

There was not a significant difference in the actual geographical distance between G (7047 km, $SD = 2673$) and NG (8203 km, $SD = 2754$) or between high (8016 km, $SD = 2455$) and low chunk strength (7234 km, $SD = 3014$) strings ($p = .186$, $\eta_p^2 = .05$ and $p = .374$, $\eta_p^2 = .02$ respectively). By contrast, there was a significant difference in geographical distance between short (5198 km, $SD = 1574$) and long-distance (9418 km, $SD = 1901$) strings, $p < .001$, $\eta_p^2 = .59$, in the expected direction. The average geographical distance for training items was 6683 km ($SD = 2776$), closer to the short-distance ($p = .059$, $\eta_p^2 = .10$) than to the long-distance test strings ($p < .001$, $\eta_p^2 = .26$), as intended. In sum, the grammaticality and similarity structures in the test phase did not conform to or violate distance expectations in themselves; rather, it is the relation of the cover task to training items that was congruent or incongruent with expectations.

2.3. Procedure and tasks

All participants were informed that the study consisted of two parts. In the first part, they had to observe sequences of cities that corresponded to the routes of an airline company. Learning was incidental, in that participants did not know that (or what) they were supposed to learn, nor that they were going to be tested.

2.3.1. Exposure phase

The two consistency conditions differed in terms of the cover stories given to participants, which favored more short-distance trips in the consistent condition and more long-distance ones in the inconsistent one (Appendix B).

Fig. 4. The mapping between letters and cities used in the present research.

The training strings of city names were individually presented on a computer screen. In the baseline learning condition, each string remained on the screen until participants pressed the space bar to move on to the next string (presentation time was limited to a maximum of 30 s). Participants were asked to press the space bar as quickly as possible. Training items were presented three times, with a different random order each time, such that no item was presented $n + 1$ times before all items had been presented n times.

In the dual task condition, we employed a random number generation (RNG) task that is known to engage executive function (Baddeley, Emslie, Kolodny, & Duncan, 1998; Towse, 1998; Towse & Cheshire, 2007). Throughout training, participants had to generate a random number between 1 and 10, guided by the beats of an electronic metronome, at the rate of 40 beats per minute. Participants were corrected if they did not keep pace with the metronome or did not produce apparently random sequences.

Under time pressure training, strings appeared for a short time that was set on the basis of string length, with strings containing seven, six, five, four and three city names appearing on the screen for 1200 ms, 1000 ms, 800 ms, 650 ms and 500 ms, respectively.

2.3.2. Test phase

After training, participants were asked to determine which of the new sequences were compatible with the old ones and which were not (Appendix B provides verbatim instructions). The test strings of city names were presented individually and remained on the screen until participants pressed either of two keys labeled "yes" or "no". Each string was presented once with no corrective feedback.

Following each response, participants had to specify the source of their knowledge by pressing one of the following keys labeled: *guess*, *intuition*, *implicit prior knowledge*, *explicit prior knowledge*, *rules*, or *memory*. Participants were given the following instructions: Press the 'guess' key when it seems to you that your response is based on no information whatsoever, that is, when you feel as if your response was decided by the flipping of a coin. Press the 'intuition' key when you are, to some extent, confident about your response and its correctness, but you have no idea why your judgment is correct. Press the 'implicit prior knowledge' key when it seems to you that your judgment was not based on any knowledge you acquired during the first part, but rather on pre-existing knowledge about cities, which, however, you could not specify and report if required. Press the 'explicit prior knowledge' key when it seems to you that your judgment was not based on any knowledge you acquired during the first part, but rather on pre-existing knowledge about cities, which you could specify and report if required. Press the 'rules' key when it seems to you that your judgment was based on a rule or some rules you learned in the first part, and which you could report if asked. Press the 'memory' category key when it seems to you that your judgment was based on memory for specific routes or parts of routes from the first part.

Table 2
Percentage of correct responses in terms of the three knowledge types (i.e., grammaticality, similarity and distance) in baseline, time pressure and dual task conditions, and whether they were significantly greater than chance (50%).

Knowledge type	Consistency condition	Training type	*M*	*SD*	*B*
Grammaticality	Consistent	Baseline	59**	9	755849
	Inconsistent		52	9	0.59
	Consistent	Time pressure	50	9	0.16
	Inconsistent		51	9	0.28
	Consistent	Dual task	56**	9	219
	Inconsistent		54*	8	11.35
Similarity	Consistent	Baseline	54*	8	11.35
	Inconsistent		56**	9	219
	Consistent	Time pressure	48	8	0.06
	Inconsistent		56**	8	1126
	Consistent	Dual task	52	8	0.66
	Inconsistent		51	8	0.27
Distance	Consistent	Baseline	58**	10	3730
	Inconsistent		57**	13	28.61
	Consistent	Time pressure	52	13	0.51
	Inconsistent		60**	10	676633
	Consistent	Dual task	50	11	0.19
	Inconsistent		57**	11	141.63

Note: B refers to a Bayes factor to test the hypothesis that there is learning (represented as a half-normal with a SD of 10% above baseline, given that 55–60% performance is typically what is achieved by participants in grammars balancing different knowledge types; e.g., Vokey & Brooks, 1992). A B of above 3 indicates strong evidence for learning and below 1/3 strong evidence for chance performance; a B between 3 and 1/3 indicates data insensitivity for deciding whether or not there was any learning (see Dienes, 2008, 2011).
** $p < .01$.
* $p < .05$.

The first three categories correspond to structural knowledge, the content of which participants claim to be unaware of, and the last three to structural knowledge, the content of which participants claim to be aware of. Thus, the first three categories were considered as indices of implicit structural knowledge, whereas the last three as indices of, at least partially, explicit structural knowledge (for use of such attributions see e.g., Dienes, 2012; Dienes, Baddeley, & Jansari, 2012; Dienes & Scott, 2005; Grey, Williams, & Rebuschat, 2014; Li, Jiang, Guo, Yang, & Dienes, 2013; Rebuschat, Hamrick, Sachs, Riestenberg, & Ziegler, 2013; Williams & Rebuschat, 2012).

2.3.3. Post experimental session

After the test phase, all participants answered a question regarding their geographical knowledge of our cities. It was noted that this question was independent of the experimental session. All 25 pairs of cities used to create the stimuli were presented, individually, on the computer screen. Participants had to provide an estimate of how long the distance between the two cities seemed to them, by pressing one of the three keys: *short*, *medium*, *long*.

3. Results

Performance in terms of the three knowledge types (grammaticality, similarity, distance) is used to assess whether learning took place. Performance on subjective measures is used to establish whether training resulted in implicit (unconscious) or explicit (conscious) knowledge of the three possible knowledge kinds.

3.1. Classification performance

Given we have employed a knowledge type (distance), different from the kind of knowledge types investigated in past AGL research, we analyzed each knowledge type separately, using two-way ANOVAs, with consistency (consistent vs. inconsistent) and training (baseline, time pressure, dual task), as between-participants variables.[2] This allows an examination of the impact of the key manipulation, regarding the consistency of knowledge, on each of the three knowledge types, and across the three training conditions.

In terms of grammaticality, a consistency by training interaction, $F(2,174) = 4.02$, $p = .020$, $\eta_p^2 = .04$, revealed that participants in the consistent condition significantly outperformed those in the inconsistent condition in the baseline learning condition only, $F(1,58) = 9.33$, $p = .003$, $\eta_p^2 = .14$, with evidence of grammaticality knowledge being found in the consistent condition, $r^2 = .49$, but not the inconsistent one $r^2 = .03$ (Table 2). This was effectively the conclusion of Pothos (2005). In the time pressure and dual task conditions, there were no significant differences in grammaticality performance between

[2] The 3-way interaction was not significant. No conclusions will be drawn about differences between the knowledge types, where not tested and analogously for related cases.

Table 3
Percentage of knowledge attributions.

Consistency condition	Training type	Guess		Intuition		Implicit prior knowledge		Explicit prior knowledge		Rules		Memory	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Consistent	Baseline	7	6	17	15	6	8	7	13	38	17	24	17
Inconsistent		13	12	15	15	5	8	3	7	39	27	25	22
Total		10	10	16	15	6	8	5	11	39	23	25	19
Consistent	Time pressure	13	11	16	13	7	11	5	8	33	25	26	16
Inconsistent		17	16	22	14	13	17	8	11	22	22	19	19
Total		15	14	19	14	10	15	7	10	27	24	22	17
Consistent	Dual task	20	11	28	14	10	10	6	9	16	17	18	16
Inconsistent		23	19	31	24	11	13	5	9	15	21	16	19
Total		21	16	30	19	10	11	5	9	17	22	17	17

the consistent and inconsistent conditions, $F(1,58) = .64$, $p = .426$, $\eta_p^2 = .01$, 95% *CI* [−6.22, 2.66] and $F(1,58) = .65$, $p = .422$, $\eta_p^2 = .01$, 95% *CI* [−2.6, 6.1], respectively. Evidence of grammaticality knowledge in the two consistency conditions collapsed together was found in the dual task condition (55%, $p < .001$, $r^2 = .23$) and not under time pressure (51%, $p = .662$, $r^2 = .003$), a difference which was significant, $F(1,116) = 6.94$, $p = .010$, $\eta_p^2 = .06$.

In terms of (chunk strength) similarity, a consistency by training interaction, $F(2,174) = 4.99$, $p = .008$, $\eta_p^2 = .05$, revealed a clear advantage of the inconsistent over the consistent condition under time pressure, $F(1,58) = 15.19$, $p < .001$, $\eta_p^2 = .21$, with only the former condition demonstrating evidence of similarity knowledge in this training type (Table 2). There was not a significant consistency effect for either baseline, $F(1,58) = 1.60$, $p = .211$, $\eta_p^2 = .03$, 95% *CI* [−7.1, 1.6], or dual task conditions, $F(1,58) = .40$, $p = .532$, $\eta_p^2 = .01$, 95% *CI* [−2.8, 5.3]. Evidence of similarity knowledge in the two consistency conditions collapsed together was found in the baseline condition (55%, $p < .001$, $r^2 = .26$), but not in the dual task one (51%, $p = .176$, $r^2 = .03$), a difference which was significant, $F(1,116) = 5.90$, $p = .017$, $\eta_p^2 = .05$.

Finally, we compare participants' performance on the distance factor. Recall, in all cases, higher performance on the distance factor was a measure of selections consistent with participants' expectations, but in the consistent condition these expectations were further consistent with the structure of the stimuli, whereas this was not the case in the inconsistent condition. There was only a significant effect of consistency, $F(1,174) = 7.38$, $p = .007$, $\eta_p^2 = .04$, with overall distance accuracy being significantly higher in the inconsistent than in the consistent condition, irrespective of training type. Evidence of distance knowledge collapsed across training types was found in both the consistent condition (54%, $p = .006$, $r^2 = .08$) and the inconsistent condition (58%, $p < .001$, $r^2 = .35$). Note that, although pre-existing biases might explain the difference between the consistent and inconsistent conditions for any one training condition, differences between training conditions cannot be so explained.

To sum up, in many cases, performance in the inconsistent condition was higher than in the consistent condition. This was the case for the distance factor, independent of training condition, and for similarity under time pressure. However, for baseline learning, grammaticality performance was higher in the consistent condition.

3.2. Subjective measures of awareness

The analysis of the conscious status of participants' structural knowledge was used to examine what sorts of information (distance relations, grammaticality or similarity) led to explicit and implicit knowledge.

3.2.1. Frequency of knowledge attributions

The frequencies of the different knowledge attributions are shown in Table 3. The guess, intuition and implicit prior knowledge categories were combined to create implicit attribution scores, and the three explicit attributions were combined to create explicit attribution scores. The two-way ANOVA on percentage of implicit attributions in the two consistency conditions and the three training conditions revealed a significant effect of training, $F(2,174) = 24.87$, $p < .001$, $\eta_p^2 = .22$. An LSD post hoc test revealed that each training type differed from the other two, *ps* < .05, with baseline learning resulting in the lowest frequency of implicit attributions (31%), dual task leading to the highest frequency (61%), and time pressure lying in between (44%).

An effect of consistency, $F(1,174) = 5.37$, $p = .022$, $\eta_p^2 = .03$, showed that, in the consistent condition, participants gave fewer implicit attributions (41%) than in the inconsistent condition (49%).

In sum, baseline learning led to the lowest and dual task to the highest number of attributions indicating implicit knowledge. In addition, the inconsistent condition led to more implicit attributions than the consistent one.

3.2.2. The implicitness and explicitness of knowledge

The accuracy of the implicit attributions (i.e., accuracy for items included in the implicit attribution scores) was assessed with three 2-way ANOVAs on participants' accuracy while providing implicit attributions for each knowledge type, with consistency (consistent vs. inconsistent) and training (baseline, time pressure, dual task), as between-participants variables. The

Table 4
Accuracy of the three performance types for implicit and explicit attributions.

Consistency Condition	Training type	% correct judgments for implicit attributions			% correct judgments for explicit attributions		
		M	*N*	*SD*	*M*	*N*	*SD*
Grammaticality							
Consistent	Baseline	55*	29	14	60**	30	12
Inconsistent		48	30	22	54	30	14
Consistent	Time pressure	45	27	19	50	30	11
Inconsistent		52	30	12	51	28	14
Consistent	Dual task	56*	30	13	57**	30	13
Inconsistent		54	29	14	56	27	26
Similarity							
Consistent	Baseline	51	29	17	56**	30	11
Inconsistent		58*	30	22	61**	30	12
Consistent	Time pressure	51	27	19	48	30	11
Inconsistent		55	30	12	62**	28	14
Consistent	Dual task	51	30	13	56*	30	13
Inconsistent		49	29	12	56	27	20
Distance							
Consistent	Baseline	56*	29	16	58**	30	14
Inconsistent		49	30	24	62**	30	15
Consistent	Time pressure	47	27	20	52	30	19
Inconsistent		59**	30	12	60**	28	14
Consistent	Dual task	49	30	10	51	30	24
Inconsistent		55*	29	11	63**	27	20

** $p < .01$.
* $p < .05$.

accuracy of the explicit attributions (i.e., accuracy for items included in the explicit attribution scores) was assessed with corresponding ANOVAs for explicit attributions (Table 4).[3]

3.2.2.1. Implicit knowledge. With respect to grammaticality, there were no significant effects of consistency, $F(2,169) = .26$, $p = .611$, $\eta_p^2 = .02$, or training type, $F(2,169) = 2.45$, $p = .089$, $\eta_p^2 = .03$, and the consistency by training interaction reached only marginal significance, $F(2,169) = 3.00$, $p = .053$, $\eta_p^2 = .03$. The overall accuracy of implicit grammaticality knowledge (52%) was close to chance, $p = .167$, $r^2 = .01$, 95% *CI* [49.3, 54.2].

No significant effects or interactions were found in terms of similarity, either, $ps > .10$, η_p^2s $< .03$. However, the overall accuracy of implicit similarity knowledge (53%) exceeded chance, $t(174) = 1.99$, $p = .048$, $r^2 = .02$.

In terms of the distance factor, there was a consistency by training interaction, $F(2,169) = 5.31$, $p = .006$, $\eta_p^2 = .06$. The consistent and inconsistent conditions did not differ in terms of the accuracy of implicit distance knowledge under the baseline condition, $F(1,57) = 1.95$, $p = .168$, $\eta_p^2 = .03$, 95% *CI* [−3.2, 18.1]. The overall accuracy of implicit distance knowledge in the two conditions collapsed together was close to chance, $p = .384$, $r^2 = .01$, 95% *CI* [47.0, 57.7]. By contrast, under time pressure, the inconsistent condition demonstrated significantly greater accuracy of implicit distance knowledge than the consistent condition, $F(1,55) = 7.63$, $p = .008$, $\eta_p^2 = .12$. In fact, evidence of implicit distance knowledge under time pressure was found only in the inconsistent, $r^2 = .39$, and not in the consistent condition, $p = .489$, $r^2 = .02$, 95% *CI* [39.4, 55.2]. Similarly, in the dual task condition, the inconsistent condition outperformed the consistent one in terms of the accuracy of implicit distance knowledge, $F(1,57) = 4.05$, $p = .049$, $\eta_p^2 = .07$, which again exceeded chance only in the former, $r^2 = .14$, and not in the latter condition, $p = .543$, $r^2 = .01$, 95% *CI* [45.1, 52.7].

In sum, there was evidence only of implicit knowledge of similarity and not of grammaticality, independent of consistency and training conditions. With respect to the distance factor, baseline learning did not elicit any implicit knowledge, independent of consistency condition. By contrast, time pressure and dual task led to the development of implicit distance knowledge, but only in the inconsistent and not in the consistent condition.

3.2.2.2. Explicit knowledge. In terms of the accuracy of explicit grammaticality knowledge, there were no significant effects or interactions, $ps > .05$, η_p^2s $< .04$. The overall accuracy of explicit grammaticality knowledge collapsed over conditions (54%) exceeded chance, $p < .001$, $r^2 = .08$.

With respect to similarity, there was a significant consistency by training interaction, $F(2,169) = 3.77$, $p = .025$, $\eta_p^2 = .04$. There was no difference between the consistent and inconsistent conditions in terms of accuracy of explicit similarity knowledge, $F(1,58) = 2.53$, $p = .117$, $\eta_p^2 = .04$, 95% *CI* [−10.8, 1.2], under baseline training, with the corresponding accuracy collapsed over the two consistency conditions (58%) exceeding chance, $p < .001$, $r^2 = .34$. Under time pressure, the inconsistent condition outperformed the consistent condition, $F(1,56) = 16.68$, $p < .001$, $\eta_p^2 = .23$, with accuracy of explicit similarity knowledge

[3] Only the 3-way interaction for the accuracy of implicit knowledge was significant ($F(4,338) = 2.58$, $p = .037$, $\eta_p{}^2 = .03$).

significantly exceeding chance only in the former, r^2 = .41, and not in the latter condition, p = .335, r^2 = .03, 95% CI [43.9, 52.1] (see Table 4). Under dual task learning, there was no significant difference in explicit similarity knowledge between the consistent and inconsistent conditions, $F(1,55) < .001$, p = .987, $\eta_p^2 < .001$, 95% CI [−8.5, 8.7]. Explicit similarity knowledge collapsed over the two consistency conditions under dual task training (56%) exceeded chance, p = .007, r^2 = .12.

In terms of distance, the inconsistent condition outperformed the consistent one, $F(1,169) = 8.52$, p = .004, η_p^2 = .05, independent of training, with accuracy of explicit distance knowledge significantly exceeding chance only in the former (62%), $p < .001$, r^2 = .34, and not in the latter condition (54%), p = .088, r^2 = .03, 95% CI [49.5, 57.7].

In sum, there was evidence of explicit grammaticality knowledge in all conditions collapsed together. There was above chance explicit similarity knowledge in the two consistency conditions, under baseline and dual task conditions. However, the inconsistent condition was the only condition that led to significant explicit similarity knowledge under time pressure. An analogous advantage of the inconsistent over the consistent condition was found in terms of explicit distance knowledge, independent of training type.

3.3. Post experimental question

The post experimental question assessed participants' knowledge of geography, by asking them to rate if trips between two cities corresponded to short, medium or long distances. Scores of 0, 1 and 2 corresponded to short, medium and long distances, respectively. The 2 × 3 × 3 (Consistency [consistent vs. inconsistent] × Training [baseline vs. time pressure vs. dual task] × Distance [short vs. medium vs. long]) ANOVA showed that only distance, $F(2,348) = 905.72$, $p < .001$, η_p^2 = .84, and its interaction with training, $F(4,348) = 3.90$, p = .004, η_p^2 = .05, were significant. The three training conditions differed only in terms of short-distance ($F(2,174) = 4.78$, p = .010, η_p^2 = .05) and long-distance items ($F(2,174) = 3.37$, p = .037, η_p^2 = .04), and not in terms of the medium ones ($F(2,174) = .54$, p = .582, η_p^2 = .01). An LSD post hoc test revealed a difference between baseline and the other two conditions in terms of the short- and long-distance items (ps < .05), with participants in the baseline condition giving lower scores for the short-distance items and higher scores for the long-distance ones than did participants in the other two training conditions. This finding does not add any useful information to our main question of interest concerning participants' ability to distinguish between the different distances, which would indicate good geography knowledge. The important finding is that the effect of distance was significant in all training conditions (baseline: $F(2,116) = 436.54$, $p < .001$, η_p^2 = .88, time pressure: $F(2,116) = 240.88$, $p < .001$, η_p^2 = .81, dual task: $F(2,116) = 261.78$, $p < .001$, η_p^2 = .82), with all three distance means differing significantly from each other and in the expected direction in all training types (all ps < .001). In particular, short-distance bigrams received the lowest score (baseline: M = .26, SD = .18, time pressure: M = .38, SD = .25, dual task: M = .35, SD = .21), long-distance bigrams received the highest score (baseline: M = 1.54, SD = .24, time pressure: M = 1.44, SD = .26, dual task: M = 1.42, SD = .32) and medium bigrams fell between these (baseline: M = 1.02, SD = .39, time pressure: M = 1.10, SD = .39, dual task: M = 1.05, SD = .36). Thus, participants had knowledge of geography in all training conditions.

4. General discussion

Our aim was to investigate how (in)congruence of prior knowledge with stimulus structure affects (a) the learning of the novel distance knowledge type involving prior knowledge relations, and (b) the learning of purely structural knowledge that has been used in past AGL research (i.e., grammaticality and similarity, as chunk strength). Pothos (2005) first employed a prior knowledge manipulation and reported that, under baseline learning conditions, when stimuli reflected only a single type of knowledge (grammaticality), consistency between prior knowledge and stimulus structure enhanced learning. We tested the generality of Pothos's (2005) findings by using different training conditions (baseline, time pressure, dual task), examining performance in a refined way, with stimuli made to reflect three knowledge types (grammaticality, similarity, distance), and including manipulations to test for the implicit/explicit nature of the knowledge acquired.

4.1. The impact of the consistent vs. inconsistent knowledge manipulation in AGL

In the baseline condition, which corresponded to that of Pothos (2005), we replicated his earlier finding, regarding grammaticality. The relevant grammaticality scores in the consistent and inconsistent conditions in Pothos (2005) were 60.2% and 52.5%, and in our work they were 59% and 52%. However, the more comprehensive approach adopted in the present study revealed ways in which Pothos's (2005) conclusion could not be generalized (in relation to other knowledge types and other learning circumstances).

Inconsistency between general knowledge and structure led to an advantage in terms of distance knowledge (independent of training type) and also in the use of similarity (under time pressure). Higher performance of the consistent condition over the inconsistent one was observed for grammaticality in the baseline learning condition (as Pothos, 2005, found too).

Conflict monitoring theory (e.g., Botvinick et al., 2001; Verguts & Notebaert, 2009) leads us to expect that the conflict between prior expectations and structural stimulus aspects in the inconsistent condition increased participants' attention to distance information, leading to greater use of distance knowledge and to greater learning of chunks that could be defined as distances.

Why should grammaticality knowledge be impaired in the inconsistent rather than consistent condition? In our materials, grammaticality was fully counterbalanced with chunk strength. Thus, plausibly, part of what constitutes grammaticality knowledge is knowledge of repetition patterns (Brooks & Vokey, 1991; Scott & Dienes, 2008); that is, which positions are repetitions of elements in which other positions (e.g., the last position being a repeat of whichever element is in the first, etc.). We determined the "global repetition structure" of each string (e.g., the letter strings MTTVMT and BAACBA both have global repetition structure 122312). Following the procedure of Scott and Dienes (2008), we determined the "global repetition proportion" of each test string, defined as the maximum proportion of a test string's global repetition structure that appeared in full (uninterrupted) in any of the training strings. This proportion was substantially higher for grammatical strings (.86, SD = .14) vs. non-grammatical strings (.66, SD = .14), $p < .001$, $\eta_p^2 = .36$; but was almost identical for high (.74, SD = .15) vs. low (.78, SD = .20) chunk strength strings, $p = .543$, $\eta_p^2 = .01$, and also for short (.80, SD = .19) vs. long (.72, SD = .17) distance strings, $p = .160$, $\eta_p^2 = .05$. Awareness of incongruency may focus attention on details and local structure, allowing small chunks to be learnt better, but global structure, like long-distance repetitions (e.g., 12341) to be learnt more poorly, as reflected in grammaticality knowledge (for the role of global vs. local structure in implicit learning see Dienes et al., 2012; Fu, Dienes, Shang, & Fu, 2013; Kiyokawa, Dienes, Tanaka, Yamada, & Crowe, 2012). We acknowledge that this is a post hoc explanation for our results, though it is testable perhaps involving methodologies especially designed for investigating the global local distinction (e.g., Kiyokawa et al., 2012).

The increased distance factor performance in the inconsistent condition does not quite comply with Heit's (1998) report, that observations incongruent with prior knowledge lead to an advantage in categorization over congruent observations, only under slow-paced learning (when more resources are available). By contrast, we found that incongruent observations led to an advantage, even under time pressure and dual task conditions. Further, this finding broadly resonates with reports that prior knowledge facilitates categorization even when only a few knowledge-related features are involved or even when categories contain information that is contradictory to prior knowledge (Kaplan & Murphy, 2000; Murphy & Kaplan, 2000).

4.2. The implicitness of prior knowledge and purely structural relations

The present research assessed the (un)consciousness of prior knowledge, as applied to AGL, by examining the (un)consciousness of knowledge for distance relations. We found evidence of both explicit (conscious) and implicit (unconscious) distance knowledge. That is, prior knowledge involves both an explicit and an implicit component. The explicit component may involve recollecting inter-city distances. The implicit component may involve intuitions about the total distance of different items or implicit memories of previously encountered related trips.

The present findings diverge from the proposal that implicit learning is a passive process leaving no room for theory-based processes (Hayes & Broadbent, 1988). Rather, our results are more consistent with the idea that implicit AGL reflects selective processes (van den Bos & Poletiek, 2009) and can be related to participants' goals (Eitam, Schul, & Hassin, 2009), domain-specific constraints (e.g., Chen et al., 2011; Leung & Williams, 2011; Rohrmeier & Cross, 2013), and motivational relevance (Eitam & Higgins, 2010). Relatedly, Tanaka, Kiyokawa, Yamada, Dienes, and Shigemasu (2008) and Kiyokawa et al. (2012) have argued that implicit learning is sensitive to selective perceptual attention and to cultural expectations, respectively. In an analogous vein, we have shown that what one attends to when learning implicitly is clearly influenced by prior knowledge.

Finally, we consider the (un)consciousness of the two purely structural factors. Similarity involved both implicit and explicit knowledge, under certain training and consistency conditions. Thus, similarity knowledge may involve both implicit and explicit knowledge. In terms of grammaticality, we found evidence only of explicit knowledge. Dienes and Longuet-Higgins (2004), Jiang et al. (2012), Kuhn and Dienes (2005), Li et al. (2013), Neil and Higham (2012), and Rebuschat and Williams (2009) obtained implicit knowledge of grammaticality independently of chunk similarity. Thus, it seems that, in general, both grammaticality and similarity knowledge can be learnt implicitly, depending on context. It is not the case that one should be regarded as implicit and the other as explicit (cf. Chang & Knowlton, 2004; Higham, 1997; Pothos & Wood, 2009).

Overall, our findings suggest that prior knowledge relevant to a learning task can involve both an explicit and an implicit component and may co-exist with both explicit and implicit knowledge of purely structural aspects of the stimuli.

4.3. Learning under time pressure and dual task conditions

One interesting conclusion from the use of time pressure in this experiment is that the effect of prior knowledge on learning grammaticality requires a long enough exposure to the training strings (i.e., more than roughly a second per string). Similarity and distance knowledge could develop within about a second per string (as the above chance similarity and distance knowledge in the time pressure inconsistent condition suggests). This finding is broadly consistent with categorization research showing that general knowledge can be applied early on during categorization (e.g., Lin & Murphy, 1997; Luhmann et al., 2006; Palmeri & Blalock, 2000).

We employed the concurrent task to explore whether the different knowledge factors relied on executive resources. Distance factor performance (in the inconsistent dual task condition) indicates that this knowledge type can be acquired with diminished executive control. No evidence of similarity knowledge was found in the dual task condition. Above chance grammaticality performance in this condition showed that it is possible to develop grammaticality knowledge with disrupted executive resources (cf. Chang & Knowlton, 2004; Higham, 1997).

We expected the time pressure and dual task to be associated mostly with implicit knowledge. These predictions were confirmed only in part. Above chance implicit distance knowledge was demonstrated only in the inconsistent condition and only in time pressure and dual task conditions. Consistently with Cleeremans and Jiménez's (2002) theory, under speeded stimulus presentation in the inconsistent condition, there was not sufficient time to develop 'quality' representations corresponding to distance knowledge, hence at least part of such knowledge was implicit. Also, the finding of implicit distance knowledge in the dual task, inconsistent condition is broadly compatible with research showing that divided attention leaves implicit knowledge intact (e.g., Dienes et al., 1995; Waldron & Ashby, 2001; Ziori & Dienes, 2008). Evidence of implicit knowledge was also found in terms of similarity (and not grammaticality), independent of training type.

The explicit knowledge results, in general, followed the pattern of classification results (except for grammaticality). Thus, again the inconsistent condition outperformed the consistent one in terms of explicit similarity knowledge (under time pressure) and explicit distance knowledge (independent of training type). So, the determining factor seemed to be the inconsistency of prior knowledge with structure, with the inconsistent condition demonstrating more explicit knowledge than the consistent one, presumably because of the greater effort the former condition required.

4.4. Conclusions

The main novel aspect of this work is that it is the only AGL study that incorporates a knowledge factor, constructed to reflect people's general knowledge and expectations and dissociated from purely structural aspects of the stimuli. More specifically, we investigated how (in)congruence of prior knowledge with structure affects the development of knowledge concerning both the relevant general knowledge (the distance factor) and purely structural information (grammaticality and similarity), and we also examined the implicitness and/or explicitness of each of the three knowledge types. Congruence of prior knowledge with structure led to an advantage only in terms of grammaticality accuracy and only under baseline training, thereby replicating Pothos's (2005)finding. However, a surprising finding is that we identified training conditions and knowledge types, such that inconsistency between expectations and stimulus structure led to a clear advantage. Finally, we have identified conditions where expectations-based knowledge may involve both explicit and implicit knowledge and may co-exist with both explicit and implicit knowledge of purely structural aspects. The complex pattern of results we identified indicates that, plausibly, the same kind of knowledge can be extracted via multiple routes. Clarifying the relevant issues (e.g., the interplay between alternative learning routes; Ashby, Alfonso-Reese, Turken, & Waldron, 1998) in AGL is an under-researched, though important direction for future research (cf. Chang & Knowlton, 2004; Pothos & Wood, 2009).

Appendix A

See Table A1.

Table A1
The distance and chunk strength factors for the test stimuli, separately for the G and NG ones.

Grammatical	Distance	Chunk strength	Non-grammatical	Distance	Chunk Strength
MSVS	2[a]	6.20[b]	MSVSSR	4[a]	5.00[b]
MVRXVS	6	6.44[b]	MVRSSV	5	6.67[b]
VXM	4[a]	1.67	MVXSSX	8	3.44
VXR	3[a]	3.67	RRXVXM	7	3.22
VXRM	5	2.40	RSRMSSR	8	2.82
VXRR	3[a]	4.00[b]	RVRXRVS	3[a]	4.73[b]
VXRRM	5	3.14	RXMSVX	6	5.22[b]
VXRRR	3[a]	3.71	RXRMMM	4[a]	2.78
VXRRRR	3[a]	3.56	SVRSX	5	4.71[b]
VXRRRRM	5	3.09	SVVSSS	2[a]	5.56[b]
VXRRRRR	3[a]	3.46	SVVSVM	5	3.33
VXSSSVS	6	6.36[b]	VRMSVS	4[a]	5.00[b]
VXSSVS	6	5.56[b]	VRRSSM	2[a]	3.56
VXSV	5	5.00[b]	VRRVX	2[a]	3.14
VXSVS	6	4.00[b]	VRRXSVX	6	5.27[b]
VXV	4[a]	3.33	VRSRXMS	7	3.46
VXVRXM	7	5.89[b]	VRSSSX	4[a]	5.22[b]
VXVRXRM	8	5.73[b]	VXXRXM	6	2.78
VXVRXVS	8	5.36[b]	XMMRXSV	8	3.36
VXVS	5	2.40	XRRVSSS	2[a]	4.64[b]

[a] Short distance.
[b] High similarity (as chunk strength).

Appendix B

B.1. Instructions

B.1.1 Training phase

Instructions in the consistent condition: "You will see a set of stimuli composed of city names. The city names correspond to the routes in Europe of an airplane company on successive days (repeated city names indicate that the plane flew to a place very near the main city). The routes were planned so as to maximize the efficiency of the trips: The secret to success for the specific company is as many trips between nearby cities as possible. Short trips are more efficient, as they require less fuel, operation and maintenance expenses. Long trips are not that efficient for this company." The instructions in the inconsistent condition were exactly the same, with the following difference: "Long trips are more efficient, as the company charges a much higher price for long trips without a significant increase in the operation and maintenance expenses of its aircrafts. Short trips are not that efficient for this company."

B.1.2 Test phase

"The airline company has found a series of rules to plan its routes. The routes that these rules allow are generally more efficient, and the company has been using them for a long time. All the routes that you saw in the first part complied with these rules. In this part, you will see more routes, some of which comply with these rules and some not. You will have to decide which comply and which do not."

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*(3), 442–481. http://dx.doi.org/10.1037/0033-295X.105.3.442.

Baddeley, A. D. (2007). *Working memory, thought and action*. Oxford: Oxford University Press.

Baddeley, A. D., & Andrade, J. (1998). Working memory and consciousness: An empirical approach. In M. Conway, S. E. Gathercole, & C. Cornoldi (Eds.), *Theories of Memory II* (pp. 1–24). Sussex: Psychology Press.

Baddeley, A. D., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *Quarterly Journal of Experimental Psychology, 51*(4), 819–852. http://dx.doi.org/10.1080/713755788.

Bailey, T. M., & Pothos, E. M. (2008). AGL StimSelect: Software for automated selection of stimuli for artificial grammar learning. *Behavior Research Methods, 40*(1), 164–176. http://dx.doi.org/10.3758/BRM.40.1.164.

Botvinick, M., Braver, T., Barch, D., Carter, C., & Cohen, J. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624–652. http://dx.doi.org/10.1037//0033-295X.I08.3.624.

Boucher, L., & Dienes, Z. (2003). Two ways of learning associations. *Cognitive Science, 27*(6), 807–842. http://dx.doi.org/10.1016/j.cogsci.2003.03.001.

Brooks, L., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General, 120*(3), 316–323. doi:10.1037/0096-3445.120.3.316.

Chang, G. Y., & Knowlton, B. J. (2004). Visual feature learning in artificial grammar classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(3), 714–722. http://dx.doi.org/10.1037/0278-7393.30.3.714.

Chen, W., Guo, X., Tang, J., Zhu, L., Yang, Z., & Dienes, Z. (2011). Unconscious structural knowledge of form-meaning connections. *Consciousness and Cognition, 20*(4), 1751–1760. http://dx.doi.org/10.1016/j.concog.2011.03.003.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioural and Brain Sciences, 36*(3), 181–204. http://dx.doi.org/10.1017/S0140525X12000477.

Cleeremans, A. (1993a). *Mechanisms of implicit learning: Connectionist models of sequence processing*. Cambridge, MA: MIT Press.

Cleeremans, A. (1993b). Attention and awareness in sequence learning. In *Proceedings of the 15th annual conference of the cognitive science society*. Hillsdale, NJ: Erlbaum.

Cleeremans, A., & Jiménez, L. (2002). Implicit learning and consciousness: A graded, dynamic perspective. In R. M. French & A. Cleeremans (Eds.), *Implicit learning and consciousness: An empirical, computational and philosophical consensus in the making* (pp. 1–40). Hove, UK: Psychology Press.

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Palgrave Macmillan.

Dienes, Z. (2011). Bayesian versus Orthodox statistics: Which side are you on? *Perspectives on Psychological Sciences, 6*(3), 274–290. http://dx.doi.org/10.1177/1745691611406920.

Dienes, Z., Altmann, G. T. M., & Gao, S. J. (1999). Mapping across domains without feedback: A neural network model of transfer of implicit knowledge. *Cognitive Science, 23*(1), 53–82. http://dx.doi.org/10.1207/s15516709cog2301_3.

Dienes, Z., Altmann, G. T. M., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*(5), 1322–1338. http://dx.doi.org/10.1037/0278-7393.21.5.1322.

Dienes, Z., Baddeley, R. J., & Jansari, A. (2012). Rapidly measuring the speed of unconscious learning: Amnesics learn quickly and happy people slowly. *PLoS ONE, 7*(3), e33400. http://dx.doi.org/10.1371/journal.pone.0033400.

Dienes, Z., & Fahey, R. (1995). The role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*(4), 848–862. http://dx.doi.org/10.1037/0278-7393.21.4.848.

Dienes, Z., & Longuet-Higgins, H. C. (2004). Can musical transformations be implicitly learned? *Cognitive Science, 28*(4), 531–558. http://dx.doi.org/10.1016/j.cogsci.2004.03.003.

Dienes, Z. (2012). Conscious versus unconscious learning of structure. In P. Rebuschat & J. Williams (Eds.), *Statistical Learning and Language Acquisition* (pp. 337–364). Mouton de Gruyter Publishers.

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychological Research, 69*(5–6), 338–351. http://dx.doi.org/10.1007/s00426-004-0208-3.

Dulany, D. E. (2003). Strategies for putting consciousness in its place. *Journal of Consciousness Studies, 10*(1), 33–43.

Dulany, D. E., Carlson, R., & Dewey, G. (1984). A case of syntactical learning and judgement: How conscious and how abstract? *Journal of Experimental Psychology: General, 113*(4), 541–555. http://dx.doi.org/10.1037/0096-3445.113.4.541.

Eitam, B., & Higgins, E. T. (2010). Motivation in mental accessibility: Relevance of a representation (ROAR) as a new framework. *Social and Personality Psychology Compass, 4*(10), 951–967. http://dx.doi.org/10.1111/j.1751-9004.2010.00309.x.

Eitam, B., Schul, Y., & Hassin, R. R. (2009). Goal relevance and artificial grammar learning. *Quarterly Journal of Experimental Psychology, 62*(2), 228–238. http://dx.doi.org/10.1080/17470210802479113.

Fu, Q., Dienes, Z., Shang, J., & Fu, X. (2013). Who learns more? Cultural differences in implicit sequence learning. *PLoS ONE, 8*(8), e71625. http://dx.doi.org/10.1371/journal.pone.0071625.

Grey, S., Williams, J. N., & Rebuschat, P. (2014). Incidental exposure and L3 learning of morphosyntax. *Studies in Second Language Acquisition, 36*, 1–35. http://dx.doi.org/10.1017/S0272263113000727.

Guo, X., Zheng, L., Zhu, L., Yang, Z., Chen, C., Zhang, L., et al (2011). Acquisition of conscious and unconscious knowledge of semantic prosody. *Consciousness and Cognition, 20*(2), 417–425. http://dx.doi.org/10.1016/j.concog.2010.06.015.

Hamrick, P., & Rebuschat, P. (2012). How implicit is statistical learning? In P. Rebuschat & J. Williams (Eds.), *Statistical learning and language acquisition* (pp. 365–382). Mouton de Gruyter Publishers.

Harris, H. D., Murphy, G. L., & Rehder, B. (2008). Prior knowledge and exemplar frequency. *Memory and Cognition, 36*(7), 1335–1350. http://dx.doi.org/10.3758/MC.36.7.1335.

Hayes, N. A., & Broadbent, D. E. (1988). Two modes of learning for interactive tasks. *Cognition, 28*(3), 249–276. http://dx.doi.org/10.1016/0010-0277(88)90015-7.

Heit, E. (1994). Models of the effects of prior knowledge on category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*(6), 1264–1282. http://dx.doi.org/10.1037/0278-7393.20.6.1264.

Heit, E. (1998). Influences of prior knowledge on selective weighting of category members. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*(3), 712–731. http://dx.doi.org/10.1037/0278-7393.24.3.712.

Higham, P. A. (1997). Dissociations of grammaticality and specific similarity effects in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*(4), 1029–1045. http://dx.doi.org/10.1037/0278-7393.23.4.1029.

Jiang, S., Zhu, L., Guo, X., Ma, W., Yang, Z., & Dienes, Z. (2012). Unconscious structural knowledge of tonal symmetry: Tang poetry redefines limits of implicit learning. *Consciousness and Cognition, 21*(1), 476–486. http://dx.doi.org/10.1016/j.concog.2011.12.009.

Kaplan, A. S., & Murphy, G. L. (2000). Category learning with minimal prior knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*(4), 829–846. http://dx.doi.org/10.1037/0278-7393.26.4.829.

Kemler Nelson, D. G. (1984). The effect of intention on what concepts are acquired. *Journal of Verbal Learning and Verbal Behavior, 23*(6), 734–759. http://dx.doi.org/10.1016/S0022-5371(84)90442-0.

Kiyokawa, S., Dienes, Z., Tanaka, D., Yamada, A., & Crowe, L. (2012). Cross cultural differences in unconscious knowledge. *Cognition, 124*(1), 16–24. http://dx.doi.org/10.1016/j.cognition.2012.03.009.

Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(1), 169–181. http://dx.doi.org/10.1037/0278-7393.22.1.169.

Kuhn, G., & Dienes, Z. (2005). Implicit learning of non-local musical rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(6), 1417–1432. http://dx.doi.org/10.1037/0278-7393.31.6.1417.

Leung, J., & Williams, J. N. (2011). Constraints on implicit learning of grammatical form-meaning connections. *Language Learning, 62*(2), 634–662. http://dx.doi.org/10.1111/j.1467-9922.2011.00637.x.

Li, F., Jiang, S., Guo, X., Yang, Z., & Dienes, Z. (2013). The nature of the memory buffer in implicit learning: Learning Chinese tonal symmetries. *Consciousness and Cognition, 22*(3), 920–930. http://dx.doi.org/10.1016/j.concog.2013.06.004.

Lin, E. L., & Murphy, G. L. (1997). Effects of background knowledge on object categorization and part detection. *Journal of Experimental Psychology: Human Perception and Performance, 23*(4), 1153–1169. http://dx.doi.org/10.1037/0096-1523.23.4.1153.

Luhmann, C. C., Ahn, W., & Palmeri, T. J. (2006). Theory-based categorization under speeded conditions. *Memory and Cognition, 34*(5), 1102–1111. http://dx.doi.org/10.3758/BF03193256.

McRae, K. (2004). Semantic memory: Some insights from feature-based connectionist attractor networks. In B. H. Ross (Ed.). *The psychology of learning and motivation* (Vol. 45, pp. 41–86). San Diego: Academic Press.

Mealor, A. D., & Dienes, Z. (2012). No-loss gambling shows the speed of the unconscious. *Consciousness and Cognition, 21*(1), 228–237. http://dx.doi.org/10.1016/j.concog.2011.12.001.

Mealor, A. D., & Dienes, Z. (2013). The speed of metacognition: Taking time to get to know one's structural knowledge. *Consciousness and Cognition, 22*(1), 123–136. http://dx.doi.org/10.1016/j.concog.2012.11.009.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness and category construction. *Cognitive Psychology, 19*(2), 242–279. http://dx.doi.org/10.1016/0010-0285(87)90012-0.

Meulemans, T., & Van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*(4), 1007–1028. http://dx.doi.org/10.1037/0278-7393.23.4.1007.

Murphy, G. L., & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*(4), 904–919. http://dx.doi.org/10.1037/0278-7393.20.4.904.

Murphy, G. L., & Kaplan, A. S. (2000). Feature distribution and background knowledge in category learning. *The Quarterly Journal of Experimental Psychology, 53A*(4), 962–982. http://dx.doi.org/10.1080/713755932.

Murphy, G. L., & Wisniewski, E. J. (1989). Feature correlations in conceptual representations. In G. Tiberghien (Ed.). *Advances in cognitive science* (Vol. 2, pp. 23–45). Chichester: Ellis Horwood.

Neil, G. J., & Higham, P. A. (2012). Implicit learning of conjunctive rule sets: An alternative to artificial grammars. *Consciousness and Cognition, 21*(3), 1393–1400. http://dx.doi.org/10.1016/j.concog.2012.07.005.

Norman, E. (2010). "The unconscious" in current psychology. *European Psychologist, 15*(3), 193–201. http://dx.doi.org/10.1027/1016-9040/a000017.

Norman, E., & Price, M. C. (2012). Social intuition as a form of implicit learning: Sequences of body movements are learned less explicitly than letter sequences. *Advances in Cognitive Psychology, 8*(2), 121–131. http://dx.doi.org/10.2478/v10053-008-0109-x.

Palmeri, T. J., & Blalock, C. (2000). The role of background knowledge in speeded perceptual categorization. *Cognition, 77*(2), B45–B57. http://dx.doi.org/10.1016/S0010-0277(00)00100-1.

Pazzani, M. J. (1991). The influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*(3), 416–432. http://dx.doi.org/10.1037/0278-7393.17.3.416.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General, 119*(3), 264–275. http://dx.doi.org/10.1037/0096-3445.119.3.264.

Pothos, E. M. (2005). Expectations about stimulus structure in implicit learning. *Memory & Cognition, 33*(1), 171–181. http://dx.doi.org/10.3758/BF03195306.

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin, 133*(2), 227–244. http://dx.doi.org/10.1037/0033-2909.133.2.227.

Pothos, E. M., Chater, N., & Ziori, E. (2006). Does stimulus appearance affect learning? *The American Journal of Psychology, 119*(2), 277–301.

Pothos, E. M., & Wood, R. L. (2009). Separate influences in learning: Evidence from artificial grammar learning with traumatic brain injury patients. *Brain Research, 1275*(12), 67–72. http://dx.doi.org/10.1016/j.brainres.2009.04.019.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*(6), 855–863. http://dx.doi.org/10.1016/S0022-5371(67)80149-x.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*(3), 219–235. http://dx.doi.org/10.1037/0096-3445.118.3.219.

Reber, A. S., & Allen, R. (1978). Analogic and abstraction strategies in synthetic grammar learning: A functional interpretation. *Cognition, 6*(3), 189–221. http://dx.doi.org/10.1016/0010-0277(78)90013-6.

Rebuschat, P., Hamrick, P., Sachs, R., Riestenberg, K., & Ziegler, N. (2013). Implicit and explicit knowledge of form-meaning connections: Evidence from subjective measures of awareness. In J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 249–269). Manoa, HI: National Foreign Language Resource Center.

Rebuschat, P., & Williams, J. (2009). Implicit learning of word order. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.

Roberts, P. L., & MacLeod, C. (1995). Representational consequences of two modes of learning. *Quarterly Journal of Experimental Psychology, 48A*(2), 296–319. http://dx.doi.org/10.1080/14640749508401392.

Rohrmeier, M., & Cross, I. (2013). Artificial grammar learning of melody is constrained by melodic inconsistency: Narmour's principles affect melodic learning. *PLoS ONE, 8*(7), e66174. http://dx.doi.org/10.1371/journal.pone.0066174.

Scott, R. B., & Dienes, Z. (2008). The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1264–1288. http://dx.doi.org/10.1037/a0012943.

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*(4), 592–608. http://dx.doi.org/10.1037/0278-7393.16.4.592.

Shanks, D. R., & Channon, S. (2002). Effects of a secondary task on "implicit" sequence learning: Learning or performance? *Psychological Research, 66*(2), 99–109. http://dx.doi.org/10.1007/s00426-001-0081-2.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17*(3), 367–447. http://dx.doi.org/10.1017/S0140525X00035032.

Shelton, J. T., Elliott, E. M., & Cowan, N. (2008). Attention and working memory: Tools for understanding consciousness. *Psyche, 14*, 1–6.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science, 210*(4468), 390–398. http://dx.doi.org/10.1126/science.210.4468.390.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3–22. http://dx.doi.org/10.1037/0033-2909.119.1.3.

Smith, J. D., & Shapiro, J. H. (1989). The occurrence of holistic categorization. *Journal of Memory and Language, 28*(4), 386–399. http://dx.doi.org/10.1016/0749-596X(89)90018-1.

Smith, E. E., & Sloman, S. A. (1994). Similarity- versus rule-based categorization. *Memory and Cognition, 22*(4), 377–386. http://dx.doi.org/10.3758/BF03200864.

Sun, R. (2000). *Duality of the mind*. Mahwah, NJ: Lawrence Erlbaum Associates.

Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science, 25*(2), 203–244. http://dx.doi.org/10.1207/s15516709cog2502_2.

Tanaka, D., Kiyokawa, S., Yamada, A., Dienes, Z., & Shigemasu, K. (2008). Role of selective attention in artificial grammar learning. *Psychonomic Bulletin and Review, 15*(6), 1154–1159. http://dx.doi.org/10.3758/PBR.15.6.1154.

Towse, J. N. (1998). On random generation and the central executive of working memory. *British Journal of Psychology, 89*(1), 77–101. http://dx.doi.org/10.1111/j.2044-8295.1998.tb02674.x.

Towse, J. N., & Cheshire, A. (2007). Random number generation and working memory. *European Journal of Cognitive Psychology, 19*(3), 374–394. http://dx.doi.org/10.1080/09541440600764570.

van den Bos, E., & Poletiek, F. (2009). Structural selection in implicit learning of artificial grammars. *Psychological Research, 74*(2), 138–151. http://dx.doi.org/10.1007/s00426-009-0227-1.

Verguts, T., & Notebaert, W. (2009). Adaptation by binding: A learning account of cognitive control. *Trends in Cognitive Science, 13*(6), 252257. http://dx.doi.org/10.1016/j.tics.2009.02.007.

Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*(2), 328–344. http://dx.doi.org/10.1037/0278-7393.18.2.328.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin and Review, 8*(1), 168–176. http://dx.doi.org/10.3758/BF03196154.

Wattenmaker, W. D., Dewey, G. L., Murphy, T. D., & Medin, D. L. (1986). Linear separability and concept learning: Context, relational properties, and concept naturalness. *Cognitive Psychology, 18*(2), 158–194. http://dx.doi.org/10.1016/0010-0285(86)90011-3.

Williams, J. N, & Rebuschat, P. (2012). Statistical learning and syntax: What can be learned and what difference does meaning make? In P. Rebuschat & J. Williams (Eds.), *Statistical learning and language acquisition* (pp. 237–264). Mouton de Gruyter Publishers.

Wisniewski, E. J. (1995). Prior knowledge and functionally relevant features in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(2), 449–468. http://dx.doi.org/10.1037//0278-7393.21.2.449.

Ziori, E., & Dienes, Z. (2008). How does prior knowledge affect implicit and explicit concept learning? *Quarterly Journal of Experimental Psychology, 61*(4), 601–624. http://dx.doi.org/10.1080/17470210701255374.