Applied Psycholinguistics **33** (2012), 829–856
doi:10.1017/S0142716411000580

# Implicit and explicit knowledge in second language acquisition

PATRICK REBUSCHAT
*University of Cambridge and Georgetown University*

JOHN N. WILLIAMS
*University of Cambridge*

Received: September 17, 2009 Accepted for publication: March 14, 2011

ADDRESS FOR CORRESPONDENCE
Patrick Rebuschat, Department of Linguistics, Georgetown University, 37th and O Streets, NW, Washington, DC 20057. E-mail: per6@georgetown.edu

ABSTRACT
Language development is frequently characterized as a process where learning proceeds implicitly, that is, incidentally and in absence of awareness of what was learned. This article reports the results of two experiments that investigated whether second language acquisition can also result in implicit knowledge. Adult learners were trained on an artificial language under incidental learning conditions and then tested by means of grammaticality judgments and subjective measures of awareness. The results indicate that incidental exposure to second language syntax can result in unconscious knowledge, which suggests that at least some of the learning in this experiment was implicit. At the same time, however, it was also found that conscious (but unverbalizable) knowledge was clearly linked to improved performance in the grammaticality judgment task.

The process of implicit learning, essentially the ability to acquire unconscious knowledge, is an elementary and ubiquitous process of human cognition (for overviews, see Cleeremans, Destrebecqz, & Boyer, 1998; Perruchet, 2008; Reber, 1993; Shanks, 2005).[1] Many essential skills, including language comprehension and production, social interaction, music perception, and intuitive decision making, are largely dependent on implicit knowledge (Berry & Dienes, 1993; Reber, 1993). The term *implicit learning* was first employed by Arthur Reber (1967) to describe a process during which subjects derive knowledge from a complex, rule-governed stimulus domain without intending to and without becoming aware of the knowledge they have acquired.[2] The term *explicit learning* is usually applied to learning scenarios in which subjects are instructed to actively look for patterns, that is, learning is intentional, a process that tends to result conscious knowledge.

The field of second language acquisition (SLA) has a long-standing interest in implicit and explicit learning (see, e.g., the contributions in Ellis, 1994; Hulstijn & Ellis, 2005; Rebuschat, in press-b). To a large degree, this interest was sparked by Krashen's (1981) proposal that learners possess two independent ways of developing knowledge of a second language (L2). According to Krashen (1981), language *acquisition* is an incidental process that results in tacit linguistic knowledge, whereas language *learning* is an intentional process that results in conscious, metalinguistic knowledge. In online speech comprehension and production, learners are thought to rely exclusively on *acquired* (or implicit) knowledge. The role of *learnt* (or explicit) knowledge is to monitor utterances for mistakes. Of importance, Krashen claimed that there is no interface between explicit and implicit knowledge. For example, explicit knowledge of a rule does not help the implicit acquisition of the same rule. For these reasons, Krashen (1981) argued that L2 teaching should focus on creating the conditions for language acquisition to take place.

Krashen's (1981) proposals and their pedagogical implications generated considerable controversy (for discussion, see Ellis, 2008). But they were also partially responsible for the increasing interest in implicit and explicit processes in L2 acquisition and use. Many studies have focused on the role of explicit knowledge (e.g., Fotos, 1993; Han & Ellis, 1998). Other studies have compared implicit and explicit learning more directly, frequently with the objective of determining the effectiveness of different instructional treatments (e.g., Alanen, 1995; de Graaff, 1997; for meta-analyses, see Norris & Ortega, 2000; Spada & Tomita, 2010).

Robinson (1995), for example, compared the SLA of English syntax under different exposure conditions. Subjects in the *implicit* condition and in the *incidental* condition were exposed to the L2 rules incidentally, that is, subjects were unaware that they were receiving training and that they would be tested afterward. The difference between the two conditions is that subjects in the former group were required to focus on the ordering of words in stimulus sentences, whereas subjects in the latter group were asked comprehension questions and thus focused on the meaning of the stimuli. Subjects in the *rule-search* condition were instructed to discover the L2 rules while receiving exposure to the training set, whereas subjects in the *instructed* condition received formal explanations of the rules in addition to rule-oriented training. At test, subjects in all conditions completed a grammaticality judgment task. Classification performance served as a measure of learning. Retrospective verbal reports were used to determine whether the acquired knowledge was conscious or not. Robinson (1995) found that, in terms of overall accuracy, the instructed group scored highest, followed by the incidental group, the rule-search group, and the implicit group. Of importance, the analyses of the verbal reports showed that large numbers of subjects in all conditions noticed rules during training and were able to report them when prompted to do so. That is, Robinson's study showed that subjects can acquire L2 syntax incidentally (as evidenced by the performance of subjects in the implicit and incidental groups) but provided no evidence that exposure resulted in unconscious knowledge.[3]

The observation that adult learners can acquire nonnative syntax without intending to is supported by several studies (e.g., Cleary & Langley, 2007; Rebuschat, 2008; Rebuschat & Williams, 2006, 2009; Robinson, 2005; Williams & Kuribara, 2008). Robinson's (2005) study, for example, includes a group of subjects that were trained on L2 syntactic rules under incidental learning conditions. Adult native speakers of Japanese were asked to memorize the meanings of a number of Samoan words and then visually presented with a total of 450 Samoan sentences.

Subjects in this group received no formal instruction of the target. The training phase required subjects to understand the meaning of each stimulus sentence and to answer a comprehension question. Learning was assessed by means of two grammaticality judgment tasks and a guided sentence production task. No measure of awareness was included. Robinson (2005) found a clear learning effect (63.8% accuracy), confirming that adult learners can acquire L2 syntactic knowledge incidentally and while processing sentences for meaning. However, because the study did not include measures of awareness, it is not clear whether learning in the incidental group resulted in unconscious knowledge.

More recently, Williams and Kuribara (2008) investigated the acquisition of L2 Japanese word order rules (head direction and scrambling) under incidental learning conditions. A semiartificial language consisting of English words and Japanese syntax (*Japlish*) was used to generate the stimulus sequences. The training set, for example, included sentences such as *Student-ga dog-ni what-o offered?*, *Vet-ga injection-o gave,* and *That sandwich-o John-ga ate*. Experimental subjects were exposed to a wide variety of sentence types by means of a plausibility judgment task. During training, subjects had to judge whether the statements made were semantically plausible or not. At test, learning was assessed by means of a grammaticality judgment task. Williams and Kuribara found that experimental subjects outperformed controls on the classification task, suggesting that the training phase produced a learning effect. No measures of awareness were included in the experimental design, so it is unclear if subjects were unaware of the knowledge they have acquired.

Cleary and Langley (2007) focused on the retention of syntactic patterns as a result of incidental exposure. Adult subjects were presented with meaningless word sequences that followed the same syntactic pattern (e.g., BEAUTIFUL TRANSPORTATION SHED TEMPORARY PLANTS, MECHANICAL CONSUMERS SUBMIT COLDER SONGS, and EFFICIENT DREAMS WRITE BETTER UMBRELLAS). After exposure, subjects were given a recognition task that required them to rate the likelihood that each test item was present in the previous part of the experiment. The testing set includes three types of sequences: *studied strings* (repetitions from training), *critical lures* (novel sequences that followed the same pattern as the studied strings), and *new strings* (new sequences that did not follow the pattern of the studied strings). Cleary and Langley (2007) found, as expected, that subjects rated studied strings highest and new strings lowest. It is interesting, however, that they also found that ratings were significantly higher for critical lures than for new strings. In other words, if a novel word sequence followed the same syntactic pattern as the training sequences, subjects were more likely to think that they had encountered it before, even if this was not actually the case. Cleary and Langley (2007) suggest that this ratings effects occurs because subjects derived an abstract representation of the syntactic structure underlying the training sentences. That is, incidental exposure to new word patterns results in knowledge that is, at least partially, independent of the items utilized in training.

Preliminary evidence that the acquisition of nonnative syntax can occur not just without intention but also result in unconscious knowledge was provided by Francis, Schmidt, Carr, and Clegg (2009). Francis et al. (2009) investigated the incidental learning of basic word order patterns by adult native speakers of

English. Subjects initially performed on an oral reading task, during which they were exposed predominantly to one type of three-word sequences, depending on which group they had been assigned to. One group was exposed mostly to noun–noun–verb strings (say, STEAK COP LEARNED), whereas the other group was exposed mainly to verb–noun–noun strings (PIERCED JANE OIL). Both of these sequence types are "nondominant" in that they do not occur in English. The arrangement of the words was random, that is, no meaningful utterances were used. Subjects were presented with three-word strings and instructed to read the three words out loud. Oral reading time for each sequence was used an indirect measure of learning. The first phase of the reading task (eight learning blocks) served to expose subjects to exemplars of the nondominant word order. The second phase (one testing block) was used to assess whether subject groups had acquired knowledge about their respective training sequences and whether this knowledge could be generalized to novel sequences with the same underlying structure. Francis et al. (2009) found that reading speed decreased as a result of exposure and that this pattern held for strings encountered during the learning blocks as well as for strings that followed the nondominant word order but had not been presented before. Their study confirms that adults can acquire new word order patterns incidentally, and that the exposure resulted in abstract, generalizable representations. It is interesting that Francis et al. (2009) also report that only few subjects were able to generate strings that followed the training pattern when prompted to do so in a grammar generation task. This suggests that the incidental exposure resulted, at least in some participants, in unconscious knowledge.

In sum, a review of the literature shows that adult subjects are able to acquire the syntactic structure of a novel language without intending to and that incidental exposure can result in abstract representations. In addition, there is also some indication that incidental exposure might result in unconscious knowledge of basic word order patterns (Francis et al., 2009). However, it is yet unclear whether there is implicit learning of L2 syntax, that is, whether incidental exposure to natural language sequences can result in unconscious knowledge.

## MEASURING IMPLICIT AND EXPLICIT KNOWLEDGE

Despite the considerable interest in the topic of implicit and explicit learning within the field of SLA, we know surprisingly little about the role that implicit learning plays in SLA. This is, in part, due to methodological reasons (Williams, 2009). Although there are several theories about the role of implicit and explicit L2 learning (e.g., Ellis, 2007; Krashen, 1981) it is difficult to adjudicate between them because of the difficulty of determining whether exposure resulted primarily in implicit or in explicit knowledge. If we intend to characterize the contribution of implicit learning to SLA, we need to be able to measure whether the acquired knowledge is implicit or explicit.

In psychology, the most commonly used criterion for disentangling these types of knowledge is awareness (or lack thereof). Implicit knowledge is unconscious knowledge that subjects are generally not aware of possessing. Explicit knowledge is conscious knowledge that subjects will be aware of possessing, although they might still not be able to verbalize it. There are several ways of measuring whether or not the acquired knowledge is conscious or unconscious (for an overview, see Rebuschat, in press-a).

### Verbal reports

One of the most common procedures for assessing whether subjects have acquired implicit or explicit knowledge is to prompt them to verbalize any rules they might have noticed while performing on the experimental tasks. This is generally done during the debriefing session, that is, these are retrospective reports (for the use of concurrent reports to assess awareness at the time of encoding, see Leow, 2000; Leow & Bowles, 2005). Knowledge is considered to be unconscious when subjects show an effect of training (e.g., above-chance performance on a grammaticality judgment task), despite being unable to describe the knowledge that underlies their performance. Several studies have provided evidence for a dissociation between task performance and verbalizable knowledge (e.g., Leung, 2007; Reber, 1967; Rebuschat, 2008; Rebuschat & Williams, 2006; Williams, 2005), and there is little doubt that exposure can result in unconscious knowledge if lack of verbalization is used as a criterion for implicitness. However, the view that knowledge is unconscious when subjects fail to verbalize the knowledge they have acquired has been criticized. First, verbal reports constitute a relatively insensitive and incomplete measure of awareness, given that it is possible to be aware of something without being able to verbally express this knowledge (Shanks & St. John, 1994: *sensitivity criterion*). Second, it is unclear whether the information assessed by verbal reports is responsible for the performance on the measure of learning (Shanks & St. John, 1994: *information criterion*).

### Subjective measures

Dienes (2004, 2008) has advocated the use of subjective measures in order to assess whether the knowledge acquired during artificial grammar learning (AGL) tasks is conscious or unconscious. One way of dissociating conscious and unconscious knowledge is to collect confidence ratings and source attributions (e.g., Dienes & Scott, 2005). This can be done, for example, by asking subjects to perform on a grammaticality judgment task and to indicate, for each judgment, how confident they were in their decision (e.g., guess, somewhat confident, very confident) and what their decision was based on (e.g., guess, intuition, memory, rule knowledge). Knowledge can be considered unconscious if subjects believe to be guessing when their classification performance is significantly above chance. Dienes, Altmann, Kwan, and Goode (1995) called this the *guessing criterion*. Knowledge is also unconscious if subjects' confidence is unrelated to their accuracy. This criterion, introduced by Chan (1992), was labeled *zero correlation criterion* by Dienes et al. (1995).

Dienes (2004) suggested that, when subjects are exposed to letter sequences in an AGL experiment, they learn about the structure of the sequences. This *structural* knowledge can consist, for example, of knowledge of whole exemplars, knowledge of fragments or knowledge of rules (e.g., *A letter sequence can start with an M or a V*). In the testing phase, subjects use their structural knowledge to construct a different type of knowledge, namely, whether the test items shared the same underlying structure as the training items (e.g., *MRVXX has the same structure as the training sequences*). Dienes labeled this *judgment* knowledge. Both forms of knowledge can be conscious or unconscious. For example, a structural representation such as *An R can be repeated several times* is only conscious if it is explicitly represented, that is, if there is a higher order thought such as *I know/think/believe that an R can repeated several times* (Rosenthal, 2005). Likewise, judgment knowledge is only conscious if there is a corresponding higher order thought (e.g., *I know/think/believe that MRVXX has the same structure as the training sequences*).

Dienes and Scott (2005) assume that conscious structural knowledge leads to conscious judgment knowledge. However, if structural knowledge is unconscious, judgment knowledge could still be either conscious or unconscious. This explains why, in the case of natural language, people can be very confident in their grammaticality decisions without knowing why. Here, structural (linguistic) knowledge is unconscious, whereas (metalinguistic) judgment knowledge is conscious. The phenomenology in this case is that of intuition, that is, knowing that a judgment is correct but not knowing why. In contrast, if structural and judgment knowledge are unconscious, the phenomenology is that of guessing. In both cases the structural knowledge acquired during training is unconscious. The experiments below exemplify how subjective measures can be employed to investigate whether incidental exposure to a new syntactic system can result in unconscious knowledge in adult learners (see also Hamrick & Rebuschat, in press; Tagarelli, Borges Mota, & Rebuschat, 2011).

In sum, the case for implicit learning depends on our criterion of implicitness and the method employed to detect the conscious or unconscious status of knowledge. If retrospective verbal reports are taken as a measure of awareness, there is little doubt that learning can result in unconscious knowledge. As has been pointed out, however, the use of verbal reports is controversial as subjects might fail to report for reasons that are unrelated to the potentially unconscious state of their knowledge. The observation of a subjective threshold by means of confidence ratings or source attributions represents an alternative approach to conceptualizing and measuring awareness. If lack of awareness is taken as performance below the subjective threshold, there is general agreement about the existence of unconscious cognitive processes operating in perception (e.g., Erdelyi, 2004; Greenwald, 1992; Kihlstrom, 1987; Merikle & Daneman, 2000) and in AGL (e.g., Dienes et al., 1995; Dienes & Scott, 2005). Yet, it remains to be established whether incidental exposure to a new linguistic system can result in implicit knowledge.

## THE PRESENT STUDY

The present study had two objectives. The first was to determine whether adult learners can acquire L2 syntax implicitly, that is, without intending to and without becoming aware of the knowledge they have acquired. As discussed above, there is little doubt that adults can inquire syntactic knowledge of a new language incidentally. However, it is unclear whether incidental learning results in unconscious knowledge. Robinson (1995) found that subjects were able to verbally describe the rule system, which suggests that learning was incidental but not implicit. The findings of Francis et al. (2009) suggest that adults might acquire implicit syntactic knowledge, but their study focused on the retention of relatively simple, meaningless three-word sequences (e.g., STEAK COP LEARNED). It thus remains to be determined whether there is implicit learning of syntax when the target system resembles the grammar of a natural language more closely.

The second objective was to establish whether adult learners can acquire L2 syntax incidentally as a result of auditory exposure in a meaning-oriented task. The incidental learning studies reviewed above all involved visual presentation of the stimulus sentences (audiovisual in Williams & Kuribara, 2008). Given that language acquisition generally proceeds as a result of exposure to auditory sequences, it is yet unclear whether adult learners can acquire L2 syntax without recourse to visual stimulus presentation.

In terms of methodology, the two experiments described below are based on the AGL paradigm. An artificial system was chosen to generate the stimulus material in order to ensure that the stimuli were novel to participants. In contrast to traditional AGL studies, however, the experiments were adapted to investigate the acquisition of natural language syntax. First, the target system employed here follows the word-order rules of a natural language (German). Second, instead of relying on a lexicon that consists of meaningless units (letters, tones, nonsense words, etc.), the lexicon in the following experiments consists of English words. Both features, the use of a natural language grammar and the use of meaningful lexical units, increased the similarity of the artificial system to natural languages and the generalizability of the findings to language acquisition outside the lab. Finally, the experiments integrate several measures of awareness, namely verbal reports, confidence ratings and source judgments.

## EXPERIMENT 1

### *Method*

*Participants.* Thirty-five native speakers of English (22 women, 13 men, $M_{age} =$ 24.3 years) were recruited to take part in this experiment and distributed into experimental ($n = 20$) and control ($n = 15$) conditions. No subject had a background in German or any other V2 language. Experimental and control groups did not differ significantly across age, gender, occupation (student vs. nonstudent), and number of languages acquired (all $p > .05$). Subjects received £5 for participating in the experiment.

*Stimulus material.* A semiartificial language, consisting of English words and German syntax, was used to generate the stimulus material for this experiment (Rebuschat, 2008; Rebuschat & Williams, 2006). In creating the stimuli, English declarative sentences were rearranged in accordance with German syntax as in the examples below (1–4). In comparison to the artificial systems commonly employed in AGL research, this system has the advantage that the grammatical complexity of natural languages is maintained and semantic information is present.

(1) Simple sentence (one-clause construction; simple predicate)
  a. English: Yesterday John bought the newspaper in the supermarket.
  b. German: Gestern kaufte John die Zeitung im Supermarkt.
  c. Stimulus: Yesterday bought John the newspaper in the supermarket.
(2) Simple sentence (one-clause construction; complex predicate)
  a. English: Yesterday John has bought the newspaper in the supermarket.
  b. German: Gestern hat John die Zeitung im Supermarkt gekauft.
  c. Stimulus: Yesterday has John the newspaper in the supermarket bought.
(3) Complex sentence (two-clause construction; sequence: main–subordinate)
  a. English: Last year Susan visited Melbourne because her daughter studied in Australia.
  b. German: Letzte Jahr besuchte Susan Melbourne, weil ihre Tochter in Australien studierte.
  c. Stimulus: Last year visited Susan Melbourne because her daughter in Australia studied.
(4) Complex sentence (two-clause construction; sequence: subordinate–main)
  a. English: Since his parents needed groceries, David purchased everything necessary.
  b. German: Weil seine Eltern Lebensmittel brauchten, kaufte David alles Notwendige ein.
  c. Stimulus: Since his parents groceries needed, purchased David everything necessary.

As is evident from the examples, the elements within phrase boundaries were left intact, whereas the specific ordering of the phrases was altered. In (1), for example, the verb phrase (VP) was moved from third position in the phrasal sequence to second. In (2), the auxiliary was placed in second position, whereas the participle was moved to the end of the sentence. In (3), the VP of the main clause was moved to second position, whereas the VP of the subordinate clause was placed in final position. Finally, in (4) the VP of the subordinate clause was moved to final position, whereas the VP of the main clause was shifted to first position.

The linguistic focus in this experiment was on four rules that determine the placement of VPs in the semiartificial language. The verb placement rules in this experiment were based on German syntax and stated that, depending on the type of predicate (simple vs. complex), the type of clause (main vs. subordinate) and the type of clause sequence (main–subordinate vs. subordinate–main), finite verbs had to be placed either in first (V1), second (V2), or final position (VF) in terms of the phrasal sequence. Table 1 illustrates the four rules in question.

Rules V2 and split VP applied to main clauses that were not preceded by a subordinate clause. They differed in that the former rule applied to simple predicates and the latter rule to complex predicates. In the semiartificial language, simple predicates occurred both in simple and in complex sentences; complex predicates only occurred in simple sentences. Rule V1 also applied to main clauses but only to those that were preceded by a subordinate clause. In contrast, rule VF applied to all subordinate clauses, irrespectively of whether a main clause preceded or followed.

A total of 192 sentences were drafted for this experiment. The sentences were read out by a male native speaker of English, digitally recorded on a Sony Mini-Disc player (MZ-R700) and subsequently edited with sound processing software

Table 1. *Descriptions and examples of the four verb placement rules*

| Rule | Description | Examples |
|---|---|---|
| V2 | Finite verb placed in second phrasal position of main clauses that are not preceded by a subordinate clause | Today bought John the newspaper in the supermarket. |
| Split VP | For complex predicates in main clauses that are not preceded by a subordinate clause: auxiliary placed in second position, participle placed in final position | Today has John the newspaper in the supermarket bought. |
| V1 | Finite verb placed in first position in main clauses that are preceded by a subordinate clause | Because his parents the newspaper in the supermarket bought, spent John the evening in his study. |
| VF | Finite verb placed in final position in all subordinate clauses | Peter repeated today that the movers his furniture scratched. |

*Note:* V, verb; VP, verb phrase.

(Audacity, version 1.2.4). The stimulus sentences were divided into a training and a testing set.

TRAINING SET. The training set consisted of 128 sentences and was subdivided into 64 plausible and 64 implausible constructions. In other words, half the sentences in the training set were syntactically correct but expressed semantically implausible propositions. Plausible and implausible items were designed so that participants would have to process the entire auditory string before being able to judge its plausibility. In most instances, the final part of the sentence would reveal whether the sentence was plausible or not, and participants had to process the entire string until reaching their judgment. The sentences below are examples of plausible and implausible constructions (5–6).

(5) Plausible constructions
  a. Chris *entertained* today his colleagues with an interesting performance. [V2]
  b. Brian *has* usually many shots during his matches *defended*. [V2VF]
  c. George *repeated* today that his students about their classes *cared*. [V2–VF]
  d. Since his teacher criticism *voiced*, *put* Chris more effort into his homework. [VF–V1]

(6) Implausible constructions
  a. ? Rose *abandoned* in the evening her cats on planet Venus. [V2]
  b. ? Sarah *has* usually in the afternoon a soup *e-mailed*. [V2VF]
  c. ? Cate *confessed* today that her horse the corridor *murdered*. [V2–VF]
  d. ? After his wife a thief *surprised*, *communicated* George with the police banana. [VF–V1]

In order to train the experimental group, 32 sentences were created for each verb placement rule. There were thus four grammatical training patterns, namely,

Table 2. *Grammatical and ungrammatical patterns used in the testing set*

| Pattern | Grammatical |
|---|---|
| V2 | Yesterday scribbled David a long letter to his family. |
| V2VF | Recently have his parents an accountant consulted. |
| V2–VF | Paul argued recently that the chairman the wrong figures displayed. |
| VF–V1 | Because his children fairy tales loved, invented John many stories. |
| | Ungrammatical |
| *VF | Recently Jim the Boston Marathon in four hours ran. |
| *V3VF | Yesterday the guitar was by David smashed. |
| *V2–V1 | Recently maintained David that abstained his father from unhealthy food. |
| *VF–VF | Because his son an instrument wanted, David with the music teacher talked. |

*Note:* The sentences are examples taken from the testing set. V, verb; VF, verb final.

V2, V2VF, V2–VF, and VF–V1, all of which occurred with equal frequency. All sentences were in the past tense.

TESTING SET. The testing set consisted of 64 new sentences and was subdivided into 32 grammatical and 32 ungrammatical items. The grammatical sentences followed the same syntactic patterns as the training sentences (V2, V2VF, V2–VF, VF–V1). The ungrammatical patterns were similar to the grammatical ones with the exception that the position of the VP was incorrect (*VF, *V3VF, *V2–V1, *VF–VF). Table 2 lists the syntactic patterns of the testing set.

With the exception of a limited number of function words, no words were repeated from the training set, making the test analogous to the transfer paradigm in AGL research (Reber, 1969). A frequency analysis of the testing set indicated that the average stimulus length was 8.8 words per sentence for grammatical items and 9.1 for ungrammatical items. There was no significant difference between training and testing sets with regard to sentence length, $t$ (31) = 0.911, $p > .05$, which suggests that length could not serve as a reliable cue to grammaticality in the testing phase.

*Procedure.*

TRAINING PHASE. Experimental participants were exposed on the semiartificial language under incidental learning conditions by means of plausibility judgments. Specifically, the training task required participants to listen to the training set on an item by item basis and to judge whether the statement made was semantically plausible or not. The experiment began with a short practice session to familiarize participants with the training task. This consisted of four practice sentences that were not repeated in the actual training phase. No mention was made that the scrambling followed the word order rules of a natural language. The training sentences were presented to each participant in random order. The entire training phase took, on average, 30 min to complete. Controls were not exposed to the artificial system.

TESTING PHASE. After training, experimental participants were informed that the word order in the previous sentences was not arbitrary but that it followed a "complex system" instead. They were then instructed to listen to 64 new scrambled sentences, only half of which would follow the same rule system as the sequences they had just been exposed to. Those sentences that did obey the rules should be endorsed as *grammatical* and those that did not rejected as *ungrammatical*. For each test sentence, participants were required to decide on its grammaticality and to report how confident they were in their judgment on a binary scale (low vs. high confidence). Accuracy in the grammaticality judgment task was used as a measure of learning. The binary confidence ratings were collected in order to assess the conscious or unconscious status of participants' metaknowledge. In the case of the control group, participants were merely told that they would listen to 64 scrambled sentences and asked to judge whether or not a sentence was grammatical.

The test sentences were presented to each participant in random order. The testing phase began with a short practice session to familiarize the participants with the new task. This consisted of four trials that were not repeated in the actual testing set. Participants were given no feedback regarding the accuracy of their grammaticality decisions. The entire testing phase took approximately 15 min to complete. At the end of the experiment, all participants were given a debriefing questionnaire that prompted them to verbalize any rule or regularity they might have noticed during the course of the experiment. Finally, the questionnaire also asked participants to supply their name, age, gender, nationality, occupation and linguistic background.

### *Results*

*Grammaticality judgments.* The analysis of the grammaticality judgments showed that the experimental group classified 54.6% ($SD$ = 12.2%) of the test items correctly and the control group 51.9% ($SD$ = 5.6%). The difference between the two groups was not significant, $t\,(33) = 0.805$, $p > .05$. Further analysis showed that neither the experimental group, $t\,(19) = 1.685$, $p > .05$, nor the control group, $t\,(14) = 1.302$, $p > .05$, performed significantly above chance. The results thus indicate that the training phase did not produce a learning effect in the experimental group.

*Confidence ratings.* The following analyses only focus on the results of the experimental group. In order to apply the binary confidence technique developed by Kunimoto, Miller, and Pashler (2001), the grammaticality judgments and the confidence ratings of the experimental group were converted into $d'$ scores. In the case of the grammaticality judgments, a positive $d'_{acc}$ value indicated a higher proportion of yes responses to grammatical stimuli and a negative $d'_{acc}$ a higher proportion of yes responses to ungrammatical stimuli. Good discrimination

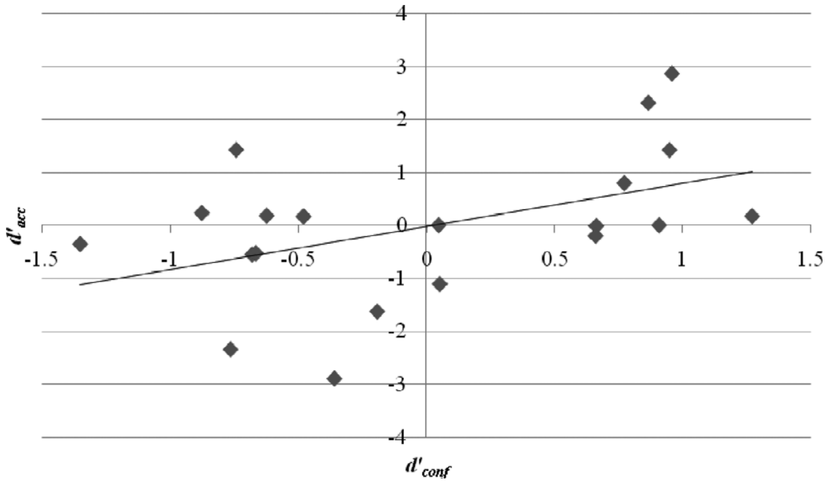


Figure 1. The $d'_{\text{conf}}$ and $d'_{\text{acc}}$ scores. Each dot represents one experimental subject.

between grammatical and ungrammatical stimuli is reflected in a positive $d'_{\text{acc}}$ value and poor discrimination in a $d'_{\text{acc}}$ value of zero or below. In the case of the confidence ratings, a positive $d'_{\text{conf}}$ indicated more correct high-confidence decisions than incorrect ones. In contrast, a negative $d'_{\text{conf}}$ suggested more incorrect high confidence decisions than correct ones. Conscious knowledge is reflected in $d'_{\text{conf}}$ values greater than zero; unconscious knowledge is reflected in $d'_{\text{conf}}$ values close to or below zero. The average $d'_{\text{conf}}$ score for the experimental group was 0.02 ($SD = 0.79$), which was not significantly different from zero, $t\ (19) = 0.116$, $p > .05$. Figure 1 plots $d'_{\text{conf}}$ against $d'_{\text{acc}}$ for all experimental subjects.

As Figure 1 indicates, 10 of the experimental subjects had a negative $d'_{\text{conf}}$ value ($M = -0.7$, $SD = 0.3$), whereas the other 10 had a positive $d'_{\text{conf}}$ value ($M = 0.7$, $SD = 0.4$), $t\ (18) = 8.752$, $p < .001$. Because a negative $d'_{\text{conf}}$ characterizes subjects in which high confidence levels are not correlated with high accuracy scores, the group with a negative $d'_{\text{conf}}$ score will be referred to as *unaware* group in subsequent analyses. In contrast, the group with the positive $d'_{\text{conf}}$ score will be referred to as *aware* group.

*Classification performance of aware, unaware, and control participants.*

OVERALL PERFORMANCE. The reanalysis of the grammaticality judgments showed that the aware group classified 60.3% ($SD$ = 10.7%) of the test items correctly and the unaware group 49% ($SD$ = 11.4%). A one-way analysis of variance (ANOVA) showed that there were significant differences between the aware, unaware, and control groups, $F\ (2, 32) = 4.371$, $p < .05$. Post hoc analysis with the Tukey honestly significant difference (HSD) test showed that the aware group significantly outperformed both the unaware group ($p < .05$) and

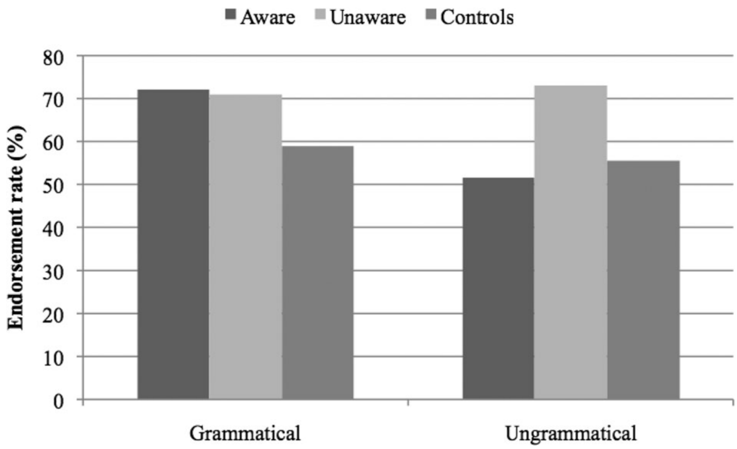


Figure 2. The classification performance of aware, unaware, and control subjects on grammatical and ungrammatical items. Bars represent endorsement rates.

the control group ($p < .05$), which suggests that the training phase did produce a learning effect but only in those participants that developed conscious knowledge. The aware group also performed significantly above chance, $t\ (9) = 3.027$, $p < .05$.

CLASSIFICATION PERFORMANCE ON GRAMMATICAL AND UNGRAMMATICAL ITEMS. The performance of the three groups on grammatical and ungrammatical items is represented in Figure 2. The results showed that the aware group endorsed 72.1% (*SD* = 16.3%) of grammatical sentences and 51.6% (*SD* = 14.9%) of ungrammatical sentences. The unaware group endorsed 70.9% (*SD* = 18.5%) of grammatical sentences and 73% (*SD* = 12.9%) of ungrammatical sentences. A one-way ANOVA further showed that there were significant differences among the three groups on both grammatical items, $F\ (2, 32) = 2.581$, $p < .05$, and ungrammatical items, $F\ (2, 32) = 6.024$, $p < .01$. Results of the post hoc analysis using the Tukey HSD test showed that aware participants significantly outperformed controls on grammatical items ($p < .05$) but not on ungrammatical ones ($p > .05$), which suggests that learning in the aware group was largely associated with the endorsement of previously encountered syntactic patterns. There were no differences between aware and unaware subjects on grammatical items, but aware subjects were significantly more likely to reject ungrammatical items ($p < .05$). Finally, there were no significant differences between unaware subjects and controls on grammatical strings, but unaware subjects were significantly more likely to endorse ungrammatical strings than the controls ($p < .05$). This suggests that subjects in the unaware group acquired some knowledge as a result of exposure, though this knowledge was largely unrelated to the actual semiartificial language. Figure 2 illustrates the performance of aware, unaware and control subjects across grammaticality.

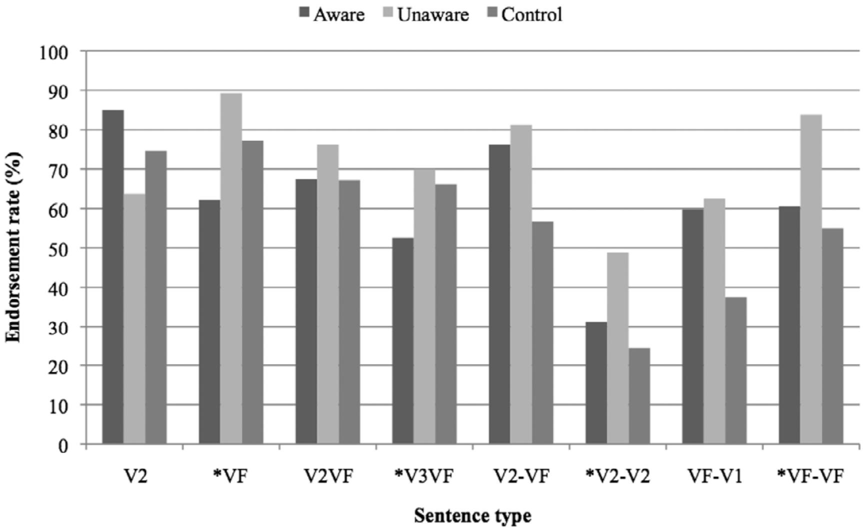


Figure 3. The classification performance of aware, unaware, and control subjects across sentence types. Bars represent endorsement rates.

CLASSIFICATION PERFORMANCE ACROSS SENTENCE TYPES. Figure 3 illustrates participants' endorsement rates across the eight syntactic patterns employed in this experiment. The scores are summarized in Table 3. As has been mentioned, the superior performance of the aware group was associated with the ability to reject ungrammatical items. Further analysis showed that there were significant differences between the three groups on V2–VF sentences, $F$ (2, 32) = 0.789, $p < .01$, and *VF–VF sentences, $F$ (2, 32) = 4.568, $p < .05$. The effects approach significance on VF–V1 sentences, $F$ (2, 32) = 3.046, $p = .06$, *VF sentences, $F$ (2, 32) = 2.704, $p = .08$, and *V2–V1 sentences, $F$ (2, 32) = 2.588, $p = .09$. The assumption of homogeneity of variance was violated for *VF–VF sentences, but the Brown–Forsythe $F$ ratio was nonetheless significant, $F$ (2, 29.9) = 5.488, $p < .05$. Results of the post hoc analyses, using the Tukey HSD test for V2–VF sentences and the Games–Howell procedure for *VF–VF sentences, showed that aware and unaware subjects were significantly more likely to endorse V2–VF sentences than the controls (both $p < .05$) and that unaware subjects were significantly more likely to erroneously endorse *VF–VF sentences than either aware or control subjects (both $p < .05$). There was no difference between aware and unaware subjects on V2–VF sentences and no difference between aware subjects and controls on *VF–VF sentences.

*Verbal reports.* The analysis of the verbal reports showed that, although most participants in this experiment were able to report knowledge, there were no verbalizations that matched the verb placement rules of the semiartificial language. No subject was able to provide descriptions of the verb placement rules employed to generate the stimulus material or to provide correct examples of grammatical

Table 3. *Mean endorsement rates (%), standard deviations, and standard error values for aware, unaware, and control participants across sentence types*

| | Sentence Type | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Groups | V2 | *VF | V2VF | *V3VF | V2–VF | *V2–V1 | VF–V1 | *VF–VF |
| Aware | | | | | | | | |
| *M* | 85* | 62.1 | 67.5† | 52.5 | 76.3* | 31.3* | 59.8 | 60.5 |
| *SD* | 31.6 | 29.3 | 25.1 | 16.5 | 18.1 | 23.8 | 30.3 | 18.7 |
| *SE* | 10 | 9.3 | 7.9 | 5.2 | 5.7 | 7.5 | 9.6 | 5.9 |
| Unaware | | | | | | | | |
| *M* | 63.8 | 89.3** | 76.3* | 70* | 81.3** | 48.8 | 62.5 | 83.8** |
| *SD* | 37 | 19.3 | 17.1 | 25.1 | 17.9 | 34.1 | 29.5 | 15.6 |
| *SE* | 11.7 | 6.1 | 5.4 | 7.9 | 5.7 | 10.8 | 9.3 | 4.9 |
| Controls | | | | | | | | |
| *M* | 74.6** | 77.3* | 67.1* | 66.1* | 56.7 | 24.4* | 37.5† | 55 |
| *SD* | 19.8 | 27.5 | 16.7 | 21.4 | 20 | 22.4 | 26.3 | 30.5 |
| *SE* | 5.1 | 7.1 | | 5.5 | 5.2 | 5.8 | | 7.9 |

*Note:* V, verb; VF, verb final.
Significance from chance: †$p < 0.1$. *$p < .05$. **$p < .001$.

sentences. The analysis of the verbal reports further suggests that participants focused exclusively on the ordering of words within clauses, disregarding important cues such as clause type and clause sequence. For example, nine subjects mentioned that verbs could occur at the end of a sentence but did not link this position to subordinate clauses. This might be taken as an indication that participants did not pay attention to phrasal arrangements above the clause level. Even though subjects were conscious of some knowledge, the analysis of the verbal reports suggests that subjects were largely unaware of the rules that determined the placement of VPs.

### *Discussion*

The results of Experiment 1 indicate that adult learners are able to acquire L2 syntax under incidental learning conditions, while processing auditory sentences for meaning, without the benefit of corrective feedback and after a relatively brief exposure period. The results further suggest that learners are able to transfer knowledge to stimuli with the same underlying structure but new surface features, which suggests that an abstract representation has been derived from the original surface form. However, the findings also show that learning of the semiartificial system was restricted to those subjects that became aware of having acquired knowledge. Although learning in the experiments was incidental, that is, a byproduct of auditory exposure to the training sentences, subjects were conscious of having acquired knowledge.

Although a learning effect was observed in the aware group, the results also showed that subjects did not acquire the relevant rule system. The acquisition of a rule system is best reflected in the capacity of dissociating grammatical and ungrammatical strings categorically. In this experiment, for example, subjects should have been able to endorse or reject a sentence purely on the basis of the placement of the verb: if the VP occupied the appropriate position, a sentence should have been endorsed; if it did not, the sentence should have been rejected. This classification performance would have resulted in high endorsement rates for grammatical stimuli and low endorsement rates for ungrammatical stimuli. The analysis of the grammaticality judgments in this experiment showed that this was not the case. The aware group performed significantly above chance on grammatical items but performed only at chance level on ungrammatical ones. The data suggests that, rather than basing decisions on rule knowledge, aware subjects relied heavily on memory for previously encountered patterns when making their grammaticality judgments. If a test stimulus matched a training pattern, as was the case with the grammatical test sequences, subjects were likely to endorse it. However, when a test stimulus did not resemble a training pattern, subjects were reduced to guessing.

On a methodological note, the results of the experiment confirm that relying on verbal reports in order to distinguish implicit and explicit knowledge is not sufficient. The verbal reports collected at the end of the experiment were helpful in determining what aspects of the semiartificial language subjects had noticed. At the same time, however, reliance on verbal reports would not have permitted the separation of the experimental group into aware and unaware subgroups. The learning effect in the aware group would thus have gone unnoticed. The binary confidence ratings, on the other hand, appeared sufficiently sensitive to detect low levels of awareness but they provided little information about what subjects noticed and were able to verbalize. The findings of Experiment 1 suggest that a battery of awareness measures, for example, confidence ratings and verbal reports, might be more adequate for the study of implicit language learning.

## EXPERIMENT 2

In Experiment 2, the focus remains on the acquisition of verb placement rules under incidental learning conditions. However, several changes were made to the methods of the previous experiment. First, only three verb placement rules determined the placement of VP in this experiment. Second, experimental subjects were trained on the grammar by means of elicited imitations and plausibility judgments. Finally, the testing phase was modified by adding a *guess* category to the confidence ratings and by adding source attributions as an additional measure of awareness. The aims of this experiment were (a) to establish whether the addition of a training task that required subjects to process word order more directly (through elicited imitations) had a positive effect on learning and (b) to evaluate the usefulness of source attributions for dissociating implicit and explicit knowledge.

### *Method*

*Participants.* Thirty native speakers of English (22 women, 8 men; $M_{\text{age}} = 24.3$ years) were recruited to take part in this experiment and evenly distributed into experimental and control conditions. The majority of participants (28) were university students and none had a background in German or any other V2 language. There were no significant differences between experimental and control groups across the variables age, gender, occupation (student vs. nonstudent), and number of languages acquired (all $p > .05$). Subjects received £5 for participating in the experiment.

*Stimulus material.* As in the previous experiment, a semiartificial language, consisting of English words and German syntax, was used to generate the stimulus material. In contrast, however, the linguistic focus was on only three verb placement rules: V2, VF, and V1. The split VP rule was removed because it was felt that the presence of two rules associated with VPs in final positions—split VP and VF—might hinder learning. The rule system in this experiment was somewhat simpler in that the placement of VP was only dependent on the type of clause (main vs. subordinate) and on the type of clause sequence (main–subordinate vs. subordinate–main). The type of predicate was no longer relevant for verb placement.

A total of 180 sentences were drafted for this experiment. The sentences were read out by a male native speaker of British English, digitally recorded on a Sony Mini-Disc player (MZ-R700) and subsequently edited with sound processing software (Audacity, version 1.2.4).

TRAINING SET. The training set consisted of 120 different sentences and was subdivided into 60 plausible constructions and 60 implausible ones. Plausible and implausible items were again designed so that participants would have to process the entire auditory string before being able to judge its plausibility. In order to train the experimental group, 40 sentences were created for each verb placement rule. That is, 40 sentences followed the V2 sentence pattern, 40 followed the V2–VF pattern, and another 40 the VF–V1 pattern. All sentences were in the past tense.

A frequency analysis of the training set showed that the average sentence length was 9.7 words per sentence in V2 constructions (9.9 for plausible items, 9.5 for implausible ones), 12.9 for VF–V1 constructions (13.2 plausible, 12.7 implausible), and 10.8 words per sentence for V2–VF constructions (10.6 plausible, 11 implausible).

TESTING SET. The testing set consisted of 60 new sentences and was subdivided into 30 grammatical and 30 ungrammatical items. The grammatical sentences followed the same syntactic patterns as the training sentences (V2, V2–VF, VF–V1). The ungrammatical templates were similar to the grammatical ones with the exception that the position of the VP was incorrect (*V1, *V3, *V4, *VF, *VF–V2, *V1–VF). With the exception of a limited number of function words, no words were repeated from the training set. Table 4 illustrates the syntactic patterns that could occur in the testing phase.

A frequency analysis of the testing set indicated that the average sentence length was 11.1 words per sentence for grammatical items and 11.6 for ungrammatical items. There was no significant difference between training and testing sets with

Table 4. *Grammatical and ungrammatical patterns used in the testing set*

| Pattern | Grammatical |
|---|---|
| V2 | Yesterday scribbled David a long letter to his family. |
| V2–VF | In the afternoon acknowledged David that her children to England moved. |
| VF–V1 | When his wife in the afternoon the office building left, prepared Jim dinner for the entire family. |
| | **Ungrammatical** |
| *V1 | Imitated Peter recently his best employee during the Christmas dinner. |
| *V3 | Recently David consulted an accountant during a five-hour meeting. |
| *V4 | In the afternoon Peter his decision undermined with poignant arguments. |
| *VF | After dinner Chloe an old car with her savings bought. |
| *VF–V2 | When her director after dinner confidential information divulged, Susan quit the department. |
| *V1–VF | Discussed Jim after dinner the new CD with his friend after his wife the kids to bed took. |

*Note:* The sentences are examples taken from the testing set. V, verb; VF, verb final.

regard to sentence length, $F(1, 193) = 0.922, p > .05$, that is, sentence length was not a reliable cue to grammaticality in the testing phase.

*Procedure.*

TRAINING PHASE. The training procedure for experimental participants was similar to the one employed in Experiment 1, with one exception. Elicited imitations were added to the training task in order to encourage subjects to process word order more directly. Participants were instructed to listen to the training set on an item by item basis, to repeat each sentence after a delayed prompt (1500 ms), and to judge whether the statement made was semantically plausible or not. The training phase took, on average, 40 min to complete. Controls did not receive any training.

TESTING PHASE. Experimental participants were informed that the word order in the previous sentences was not arbitrary but that it followed a "complex system" instead. They were then instructed to listen to 60 new scrambled sentences, only half of which would follow the same rule system as the sequences they had just been exposed to. In contrast to the previous experiment, subjects were asked to judge the basis of their grammaticality decision as well. That is, for each test sentence, participants were required to decide on its grammaticality, to report how confident they were in their judgment, and to indicate what the basis of their judgment was.

Participants could indicate their levels of confidence by selecting one of three response options: *guess*, *somewhat confident,* or *very confident*. Participants were instructed to select the *guess* category only if they had no confidence whatsoever in their classification decision and believed to be guessing. If they had a small amount of confidence, they were asked to select the *somewhat confident* category instead. If they were very confident, they should opt for the *very confident* category. In the case of the source attributions, participants were asked to select one of four response options: *guess*, *intuition*, *memory,* and *rule knowledge*. Participants were instructed to only use the *guess* category when decisions were based on real guesses, that is, they might as well have flipped a coin. The *intuition* category should be selected if participants were somewhat confident in their decisions but did not know why they were right. The *memory* category should be selected when judgments were based on the recollection of parts or entire sentences from the training phase. Finally, the *rule knowledge* category should be selected following decisions that were based on a rule that was acquired during the training phase and that subjects would be able to report at the end of the experiment. All participants were provided with these instructions before starting the testing phase.

The test sentences were presented to each participant in random order. The testing phase began with a short practice session, which consisted of four practice trials that were not repeated in the actual testing set. The entire testing phase took approximately 15 min. Afterward, participants completed the same debriefing questionnaire used in the previous experiment.

### *Results*

Performance on the grammaticality judgment task served as the measure of learning. Verbal reports, confidence ratings, and source attributions were employed to determine to what extent subjects were aware of having acquired knowledge and whether the acquired knowledge was conscious or not.

#### *Grammaticality judgments.*

OVERALL PERFORMANCE. The analysis of the grammaticality judgments showed that the experimental group classified 61.6% (*SD* = 8.3%) of the test items correctly and the control group 42.9% (*SD* = 5.1%). The difference between the two groups was significant, $t\,(27) = 7.289$, $p < .001$. Further analysis showed that the experimental group performed significantly above chance on this task, $t\,(13) = 5.150$, $p < .001$, whereas the controls scored significantly below chance, $t\,(14) = -5.361$, $p < .001$. The training phase produced a clear learning effect in the experimental participants.

In order to establish whether any improvements in accuracy occurred during the testing phase, the grammaticality judgment data was equally divided into 10 chronological stages. The mean accuracy for each block of six judgments was then computed for both groups. A repeated-measures ANOVA showed no significant differences across the stages in the experimental group, $F\,(9, 117) = 0.760$, $p > .05$, or in the control group, $F\,(5.1, 71.9) = 0.722$, $p > .05$, with Greenhouse–Geisser correction. No online learning effect was observed during the testing phase, that is, the learning effect in the experimental group was a result of exposure during the training phase.

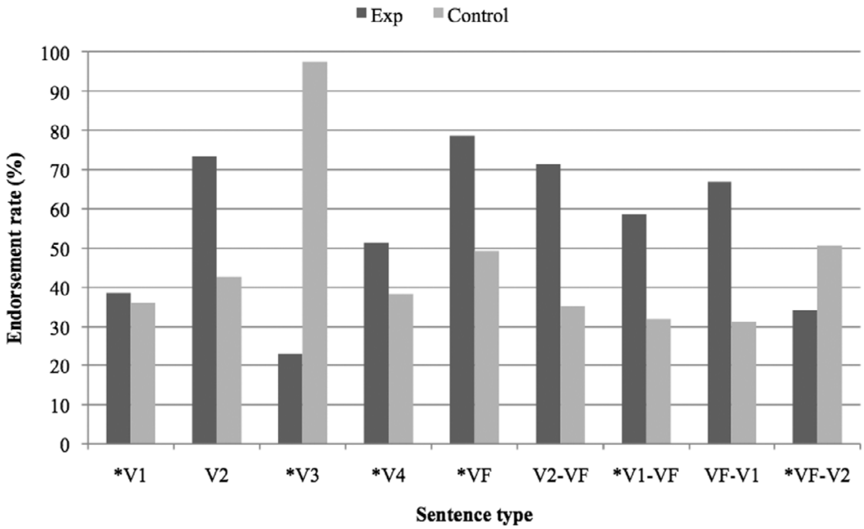


Figure 4. The classification performance of experimental and control subjects across sentence types.

CLASSIFICATION PERFORMANCE ON GRAMMATICAL AND UNGRAMMATICAL ITEMS. The following analyses report endorsement rates, rather than accuracy scores. The results showed that the experimental group endorsed 71% (*SD* = 13.5%) of all grammatical sentences and 47% (*SD* = 19.7%) of all ungrammatical ones. No experimental subject endorsed more than 90% of grammatical sequences and less than 10% of ungrammatical ones, that is, there was no evidence of categorical classification performance. The control group only endorsed 36.4% (*SD* = 30.3%) of grammatical sentences and 51% (*SD* = 28.4%) of ungrammatical ones, $t$ (14) = −5.268, $p < .001$. The difference between experimental and control subjects on grammatical items was significant, $F$ (1, 27) = 14.824, $p < .001$, that is, the experimental group was significantly more likely to correctly endorse grammatical strings. The difference between groups on ungrammatical strings was not significant, however, $F$ (1, 27) = 0.125, $p > .05$. That is, the classification performance of the experimental groups was largely driven by the correct endorsement of grammatical items.

CLASSIFICATION PERFORMANCE ACROSS SENTENCE TYPES. Figure 4 illustrates participants' endorsement rates across the nine syntactic templates employed in this experiment. The scores are summarized in Table 5. The analysis indicated that the experimental group was significantly more likely than the control group to correctly endorse all grammatical sentence types, that is, V2 sentences, $F$ (1, 27) = 6.426, $p < .05$, V2–VF sentences, $F$ (1, 27) = 15.778, $p < .001$, and VF–V1 sentences, $F$ (1, 27) = 8.831, $p < .05$. The experimental group thus displayed a clear preference for endorsing previously encountered sentence patterns. The experimental group was also significantly more likely to correctly reject *V3

Table 5. *Mean endorsement rates (%), standard deviations, and standard error values for experimental and control participants across sentence types*

| | Sentence Types | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Group | *V1 | V2 | *V3 | *V4 | *VF | V2–VF | *V1–VF | VF–V1 | *VF–V2 |
| Exp. | | | | | | | | | |
| *M* | 38.6 | 73.4* | 22.9* | 51.4 | 78.6** | 71.4** | 58.6 | 66.9* | 34.3 |
| *SD* | 36.3 | 27.6 | 35.8 | 33 | 22.8 | 18.8 | 35.5 | 29.2 | 30.8 |
| *SE* | 9.7 | 7.4 | 9.6 | 8.8 | 6.1 | 5 | 9.5 | 7.8 | 8.2 |
| Controls | | | | | | | | | |
| *M* | 36 | 42.7 | 97.3** | 38.3 | 49.3 | 35.3† | 32† | 31.3 | 50.7 |
| *SD* | 34 | 36.7 | 7 | 37.0 | 42.7 | 28.8 | 35.3 | 34.8 | 38.4 |
| *SE* | 8.8 | 9.5 | 1.8 | 9.6 | 11.0 | 7.4 | 9.1 | 9 | 9.9 |

*Note:* V, verb; VF, verb final.
Significance from chance: †$p < .07$. *$p < .05$. **$p < .001$.

sentences, $F(1, 27) = 62.401, p < .001$. That experimental subjects only endorsed 22.9% of *V3 sentences, compared to the 97.3% endorsement rate in the control group, is interesting because the latter are the most English-like structures in the experiment. It suggests that the experimental group had learned to reject a sentence type that would otherwise be acceptable in English. Experimental subjects were also significantly more likely to erroneously endorse stand-alone *VF sentences, $F(1, 27) = 5.18, p < .05$, which suggests that they had not learned that clause-final VPs were restricted to subordinate clauses. In the case of *V1–VF sentences, the difference between experimental and control subjects approaches significance, $F(1, 27) = 4.083, p = .053$. There were no significant differences between the groups on *V1, *V4, and *VF–V2.

*Confidence ratings.* The following analyses only report the results of the experimental group. The average confidence level in the experimental group was 5.9 ($SD = 1.8$). In terms of proportion, experimental participants tended to select the option somewhat confident most frequently and the guess option least frequently. In terms of accuracy, the analysis indicated that experimental participants were most accurate when reporting to be very confident in their decisions and slightly less accurate when reporting to be somewhat confident. Accuracy was lowest for those grammaticality decisions in which subjects had no confidence whatsoever. Experimental participants scored significantly above chance when reporting to be somewhat confident and very confident. When participants reported to be guessing, performance was indistinguishable from chance. The guessing criterion for unconscious judgment knowledge was thus not satisfied. Table 6 summarizes the findings.

The Chan difference score was computed in order to establish whether learning in the experimental group was implicit by the zero correlation criterion. The average confidence for correct grammaticality decisions was 6.1 ($SD = 1.6$) and the average confidence for incorrect decisions was 5.6 ($SD = 1.7$), that is, experimental

Table 6. *Accuracy and proportions (%) across confidence ratings*

| | Guess | Somewhat Confident | Very Confident |
|---|---|---|---|
| Accuracy | 53 | 60* | 65* |
| Proportion | 12 | 54 | 34 |

Significance from chance: $^*p < .05$.

Table 7. *Accuracy and proportions (%) across source attributions*

| | Guess | Intuition | Memory | Rule |
|---|---|---|---|---|
| Accuracy | 56 | 59* | 57 | 65** |
| Proportion | 10 | 32 | 15 | 43 |

Significance from chance: $^*p < .05$. $^{**}p < .001$.

participants were more confident in correct decisions than in incorrect ones. The difference (0.5) was significant, $t(13) = 2.310, p < .05$. That is, there was conscious judgment knowledge according to the zero correlation criterion. Participants were partially aware that they had acquired some knowledge during the training phase.

*Source attributions.* In terms of proportion, experimental participants most frequently believed their classification decisions to be based on rule knowledge, followed by intuition and memory (Table 7). The guess category was selected least frequently. In terms of accuracy, experimental participants scored highest when reporting to use rule knowledge to guide their decisions, followed by the intuition and memory categories. Subjects were least accurate when basing decisions on guesses (56%). Participants performed significantly above chance when basing their decisions on rule knowledge or on intuition. The latter suggests that participants acquired some unconscious structural knowledge.

*Verbal reports.* The analysis of the verbal reports showed that there were no verbalizations that matched the rules of the semiartificial language. As in Experiment 1, several subjects mentioned that verbs could occur in sentence-final position but did not associate this verb placement option with subordinate clauses. Even though subjects were conscious of some knowledge, as suggested by the analysis of the source judgments, the analysis of the verbal reports suggests that subjects were largely unaware of the rules that determine the placement of VPs.

## DISCUSSION

Experiment 2 confirms that adult learners are able to acquire the syntax of a novel language without intending to, while processing auditory sentences for meaning, without feedback and after limited exposure. Because subjects were able to transfer their knowledge to judge the grammaticality of novel stimuli, the findings also support the view that learners derived an abstract representation of the training sentences. Of importance, Experiment 2 provided evidence that incidental exposure to L2 syntax can result in unconscious knowledge. That is, at least some of the learning in this experiment was implicit (see also Hamrick & Rebuschat, in press; Tagarelli et al., 2011). The analysis of confidence ratings and source attributions showed that, although subjects were aware that they had acquired knowledge, they were at least partially unaware of what knowledge they had acquired. When attributing grammaticality judgments to intuition, subjects performed significantly above chance, that is, they had acquired unconscious structural knowledge. At the same time, it is important to highlight that subjects in Experiment 2 were significantly more accurate when reporting higher levels of confidence and when basing their decisions on explicit categories (memory and rule knowledge). Conscious, but unverbalizable knowledge was clearly linked to improved performance in the grammaticality judgment task.

Experiment 2 introduced several modifications to the design of the previous experiment. Three, rather than four, verb placement rules determined the placement of VP, which meant that the position of the verb was only dependent on clause type and on clause sequence. Elicited imitations were introduced as an additional training task, and source attributions were added to the testing phase as an additional measure of awareness. As mentioned above, Experiment 2 was designed in order to establish whether the addition of a training task that required subjects to process word order more directly had a positive effect on learning and to evaluate the usefulness of source attributions for distinguishing implicit and explicit knowledge. The results showed that the modifications produced a greater learning effect in the experimental group (61.6%, compared to 54.6% in Experiment 1). As in Experiment 1, performance was largely driven by the above-chance endorsement of grammatical sentences. Experimental participants endorsed 71% of all grammatical items but also 47% of ungrammatical items, which indicates that classification performance might have been partially based on memory for previously encountered patterns (see also Ellis et al., 2009). That classification performance was not categorical, but probabilistic, further suggests that subjects did not acquire the verb placement rules. Accuracy for grammatical items was 77.2%, but at chance (53.8%) for ungrammatical items. When participants based their classifications on rule knowledge, these rules did not correspond to the word order rules of the semiartificial language. The analysis of the verbal reports supports this view.

The performance across sentence types suggests that subjects did not learn that clause type and clause sequence determined the placement of VPs. *VF sentences, for example, were overendorsed, which means that subjects did not know that a verb could only occur at the end of a specific clause type. Subjects were more likely to endorse VF–V1 sentences than their ungrammatical counterparts, *VF–V2, but the preference for VF–V1 could be explained by the acquisition of a microrule like “a verb can follow a verb in the middle of a sentence.” There would be no need to assume that subjects had acquired the V1 rule. That subjects were just as likely to endorse V2–VF sentences as *V1–VF sentences supports the notion that subjects did not learn that clause sequence was a relevant cue to verb placement.

Because several changes were made to the previous design, it is difficult to judge what particular modification has improved performance in this experiment. One possibility is that this was due to the addition of the elicited imitations. In contrast to Experiment 1, subjects in the present experiment had to focus both on the arrangement of words and on sentence meaning. That subjects had to recall word order in order to perform on the training task could have increased learning by making them process serial order more directly. A different possibility is that the removal of the split VP rule simplified the grammar and made the structure of the stimulus sentences easier to acquire. Although subjects did not learn to associate verb positions with specific types of clauses or clause arrangements, this simplification could have led them to learn what positions were licensed by the grammar. This seems unlikely, however, given that trained subjects in Experiment 1 appear to have learned similar things. Both groups knew, for example, that a verb could come in second or in final position but not that clause type or clause sequence played a role in verb placement. A third possibility is that the ungrammatical templates used in Experiment 2 were easier to reject than those employed in the previous experiments. Experimental subjects in Experiments 1 and 2 were equally likely to endorse grammatical sentences, but subjects in the former experiment were more likely to wrongly endorse ungrammatical ones. Perhaps subjects in Experiment 2 would have performed equally poorly had they been asked to judge *V3VF, *V2–V1, and *VF–VF sentences instead of *V1, *V3, *V4, *VF, *V1–VF, and *VF–V2.

In terms of methodology, Experiment 2 further confirms that the sole reliance on verbal reports is clearly inadequate for assessing awareness. The analysis of the verbal reports showed that participants were unable to verbally describe the rules of the semiartificial system, which would have supported the erroneous assumption that learning in Experiment 2 resulted entirely in an implicit knowledge base. That subjects developed conscious judgment knowledge would have gone unnoticed. The combined use of confidence ratings and source attributions appears to be a promising method for distinguishing implicit and explicit knowledge (see also Rebuschat, in press-a) in language acquisition research.

## CONCLUSION

The results of the two experiments described above show that adult learners can acquire the syntax of a novel language without intending to and while processing auditory sequences for meaning. The findings further show that incidental exposure to the stimulus sentences resulted in an abstract, generalizable representation, and that subjects did not acquire linguistic rules. The analysis of the different awareness measures suggests that subjects were aware of having acquired knowledge and that exposure resulted in both conscious and unconscious knowledge. Although adults seem capable of implicit learning, it was also clear that conscious, but unverbalizable knowledge was linked to improved performance in the grammaticality judgment task.

The findings reported in this study have several implications for our understanding of language acquisition. Although the precise form of the knowledge acquired in these experiments is unclear, the findings provided no evidence for rule learning in the vast majority of subjects. It suggests that subjects in these types of experiments (and perhaps in natural language acquisition) do not acquire linguistic rules. The results support Shanks (1995; Johnstone & Shanks, 2001), who argues against the possibility of implicit rule learning. Additional research is necessary to characterize the nature of what was learned incidentally and to determine more precisely what conditions might lead to successful rule acquisition. For example, it would be important to establish whether increased exposure would lead to the development of rule knowledge under incidental learning conditions.

Although adults can acquire unconscious knowledge, the experiments reported in this article demonstrate that adult syntactic learning results predominantly in a conscious (but largely unverbalizable) knowledge base. This might explain why learning was very much constrained across all experiments. *VF sentences, for example, were generally accepted as grammatical, even though this option was restricted to subordinate clauses. It would be of interest to run the same, or a slightly adapted, version of Experiments 1 and 2 with children in order to determine whether there are child–adult differences in syntactic learning. Young learners, especially preschoolers without extensive metalinguistic knowledge, might display more implicit learning than adults. It would be also interesting to establish whether this would lead to greater learning effects across patterns. If this were the case, then the fact that adults are potentially less likely to engage in implicit learning of a novel language might help explain why they frequently fail to achieve nativelike levels of proficiency in a novel language, despite prolonged periods of exposure.

From a methodological perspective, the results of the experiments confirm that relying on verbal reports as a measure of awareness is not sufficient (for alternative measures, see Rebuschat, 2011). The verbal reports collected at the end of the experiment were helpful in determining what aspects of the semiartificial language subjects had consciously noticed. At the same time, verbal reports were clearly not sensitive enough to assess whether subjects were aware of the knowledge they had acquired. Confidence ratings and source attributions provided a very useful method for capturing low levels of awareness and to observe the conscious status of both structural and judgment knowledge. Future experiments on first language and L2 acquisition would benefit from the introduction of this relatively simple, but effective way of measuring implicit and explicit knowledge.

As far as the assessment of learning effects is concerned, it would be useful to explore whether the binary grammaticality judgments used in this study are an adequate measure of learning. It could be, for example, that learning would not have appeared as constrained if more sensitive measures had been used. Scott and Dienes (2008) have shown that familiarity is the essential source of knowledge in AGL, which suggests that, in terms of offline measures, familiarity or preference judgments might be more suitable. The use of online measures in particular, for example, tracking eye movements, recording response latency (e.g., in a rapid serial visual presentation task), or measuring event-related potentials (e.g., Tokowicz & MacWhinney, 2006), seems to be a promising way to assess the knowledge underlying native-speaker intuitions. Finally, it should also be noted that the two experiments above focused on comprehension tasks. Given that communication involves both comprehension and production of language, it would be important to determine whether the knowledge acquired in artificial language experiments transfers successfully to production tasks. However, further research is necessary to address the shortcomings and the unanswered questions of the present study.

## ACKNOWLEDGMENTS

We thank Zoltán Dienes, Ron Leow, and Brechtje Post for their comments on earlier drafts of this article. We are also grateful to Phillip Hamrick, Kaitlyn Tagarelli, Katie Jeong-eun Kim, and Julie Lake for helpful discussions and feedback. Finally, we thank the anonymous reviewers for their constructive suggestions.

## NOTES

1. In this paper, we will use the following definitions. Implicit learning is the learning process that results primarily in unconscious knowledge, that is, knowledge that is tacit and inaccessible to conscious introspection. Implicit learning generally occurs without the intention to learn and without awareness of what has been learned, that is, learners are often unaware of having acquired knowledge. (In keeping with the psychological literature, implicit learning will not be used to refer to "learning without attention" or "learning without noticing.") Explicit learning is the learning process that results primarily in conscious knowledge. This is usually associated with intentional learning conditions. We will use the terms *implicit knowledge* and *unconscious knowledge* as well as *explicit knowledge* and *conscious knowledge* interchangeably.
2. As Rebuschat and Williams (in press) point out, these two dimensions, *intentionality* and *awareness*, are central to the notions of implicit and explicit learning in the psychological literature. In the research tradition started by Reber (1967), the use of the term *implicit* is generally restricted to those situations where subjects have acquired unconscious knowledge under incidental learning conditions. If incidental exposure in an experiment results in conscious knowledge, for example, when subjects were able to figure out the rule system despite not having been told about its existence, the learning process is usually only characterized as being *incidental* and not as implicit. The same applies for those experiments that do not include a measure of awareness. Of importance, the three measures of awareness discussed in this paper assess the conscious or unconscious status of the acquired knowledge. That is, the focus here is on the product of learning, and not on awareness (or lack thereof) at the time of encoding.
3. Another important lesson from Robinson (1995) is that we need to be careful with wording. Subjects in the "implicit" group clearly acquired conscious knowledge, so learning in this condition was not implicit at all.

## REFERENCES

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 259–302). Honolulu: University of Hawai'i Press.

Berry, D. C., & Dienes, Z. (1993). *Implicit learning: Theoretical and empirical issues*. Hove: Erlbaum.
Chan, C. (1992). *Implicit cognitive processes: Theoretical issues and applications in computer systems design*. Unpublished doctoral thesis, University of Oxford.
Cleary, A. M., & Langley, M. M. (2007). Retention of the structure underlying sentences. *Language and Cognitive Processes, 22*, 614–628.
Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences, 2*, 406–416.
de Graaff, R. (1997). The *eXperanto* experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition, 19*, 249–276.
Dienes, Z. (2004). Assumptions of subjective measures of unconscious mental states: Higher order thoughts and bias. *Journal of Consciousness Studies, 11*, 25–45.
Dienes, Z. (2008). Subjective measures of unconscious knowledge. *Progress in Brain Research, 168*, 49–64.
Dienes, Z., Altmann, G., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1322–1338.
Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychological Research, 69*, 338–351.
Ellis, N. C. (Ed.). (1994). *Implicit and explicit learning of languages*. London: Academic Press.
Ellis, N. C. (2007). The weak-interface, consciousness, and form-focussed instruction: Mind the Doors. In S. Fotos & H. Nassaji (Eds.), *Form-focused instruction and teacher education: Studies in honour of Rod Ellis* (pp. 17–33), Oxford: Oxford University Press.
Ellis, R. (2008). Explicit knowledge and second language learning and pedagogy. In J. Cenoz & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 6. Knowledge about language* (pp. 143–153). New York: Springer Academic.
Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Bristol: Multilingual Matters.
Erdelyi, M. (2004). Subliminal perception and its cognates: Theory, indeterminacy, and time. *Consciousness and Cognition, 13*, 73–91.
Fotos, S. (1993). Consciousness-raising and noticing through focus on form: Grammar task performance versus formal instruction. *Applied Linguistics, 14*, 385–407.
Francis, A. P., Schmidt, G. L., Carr, T. H., & Clegg, B. A. (2009). Incidental learning of abstract rules for non-dominant word orders. *Psychological Research, 73*, 60–74.
Greenwald, A. G. (1992). New look 3: Unconscious cognition reclaimed. *American Psychologist, 47*, 766–779.
Hamrick, P., & Rebuschat, P. (in press). How implicit is statistical learning? In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition*. Berlin: Mouton de Gruyter.
Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research, 21*, 1–23.
Hulstijn, J. H., & Ellis, R. (Eds.). (2005). Implicit and explicit second-language learning [Special issue]. *Studies in Second Language Acquisition, 27*(2).
Johnstone, T., & Shanks, D. R. (2001). Abstractionist and processing accounts of implicit learning. *Cognitive Psychology, 42*, 61–112.
Kihlstrom, J. F. (1987). The cognitive unconscious. *Science, 237*, 1445–1452.
Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.
Kunimoto, C., Miller, J., & Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Consciousness and Cognition, 10*, 294–340.
Leow, R. P. (2000). A study of the role of awareness in foreign language behaviour: Aware versus unaware learners. *Studies in Second Language Acquisition, 22*, 557–584.
Leow, R. P., & Bowles, M. (2005). Attention and awareness in SLA. In C. Sanz (Ed.), *Mind and context in adult second language acquisition* (pp. 179–203). Washington, DC: Georgetown University Press.
Leung, J. (2007) *Implicit learning of form-meaning connections*. Unpublished doctoral dissertation, University of Cambridge.
Merikle, P. M., & Daneman, M. (2000). Conscious vs. unconscious perception. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 1295–1303). Cambridge, MA: MIT Press.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*, 417–528.

Perruchet, P. (2008). Implicit learning. In J. Byrne (Ed.), *Learning and memory: A comprehensive reference: Vol. 2. Cognitive psychology of memory* (pp. 597–621). Oxford: Elsevier.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 317–327.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology, 81*, 115–119.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford: Oxford University Press.

Rebuschat, P. (2008). *Implicit learning of natural language syntax*. Unpublished dissertation, University of Cambridge.

Rebuschat, P. (in press-a). Measuring implicit and explicit knowledge in second language research: A review. *Language Learning*.

Rebuschat, P. (Ed.). (in press-b). *Implicit and explicit learning of languages*. Amsterdam: John Benjamins.

Rebuschat, P., & Williams, J. N. (2006). Dissociating implicit and explicit learning of natural language syntax. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (p. 2594). Mahwah, NJ: Erlbaum.

Rebuschat, P., & Williams, J. N. (in press). Implicit learning in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (Vol. 20). Oxford: Wiley–Blackwell.

Robinson, P. (1995). Aptitude, awareness and the fundamental similarity of implicit and explicit second language learning. In R. Schmidt (Ed.): *Attention and awareness in foreign language learning* (pp. 303–358). Manoa, HI: University of Hawai'i Press.

Robinson, P. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, & Hernstadt (1991) and Knowlton & Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition, 27*, 235–268.

Rosenthal, D. M. (2005). *Consciousness and mind*. Oxford: Oxford University Press.

Scott, R., & Dienes, Z. (2008). The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1264–1288.

Shanks, D. R. (1995). *The psychology of associative learning*. Cambridge: Cambridge University Press.

Shanks, D. R. (2005). Implicit learning. In K. Lambert & R. Goldstone (Eds.), *Handbook of cognition* (pp. 202–220). London: Sage.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning system. *Behavioral and Brain Sciences, 17*, 367–447.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60*, 263–308.

Tagarelli, K. M., Borges Mota, M., & Rebuschat, P. (2011). The role of working memory in the implicit and explicit language learning. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2061–2066). Austin, TX: Cognitive Science Society.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition, 27*, 173–204.

Williams, J. N. (2009). Implicit learning in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 319–353). Bingley: Emerald Press.

Williams, J. N., & Kuribara, C. (2008). Comparing a nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling. *Lingua, 118*, 522–553.