15

NFLRC monographs

Implicit and Explicit Knowledge of Form-Meaning Connections: Evidence From Subjective Measures of Awareness

Patrick Rebuschat
Lancaster University

Phillip Hamrick
Rebecca Sachs
Kate Riestenberg
Nicole Ziegler
Georgetown University

Two recent studies on the possibility of learning form-meaning connections without awareness (Hama & Leow, 2010; Williams, 2005) reached contradictory conclusions. This conceptual replication and extension clarifies the differences in their results by adding subjective measures of awareness, namely confidence ratings and source attributions (Rebuschat, 2008). Experimental participants were exposed to sentences of a semi-artificial language consisting of English words and four artificial determiners (gi, ro, ul, ne). Participants were told the determiners encoded distance (near/far) but were not told about a hidden regularity involving animacy. Trained control subjects were exposed to the same sentences and instructions, but determiner animacy was randomized. On a posttest with new sentences, participants had to choose determiners from two options differing only in animacy. In addition, they also had to indicate the basis of each test response (guess, intuition, memory, rule) and their confidence (on a 4-point scale), allowing us to assess the conscious or unconscious status of their structural knowledge (of content) and judgment knowledge (knowing that they knew). Our results showed that the experimental group significantly outperformed the trained controls in terms of overall accuracy. The analysis of the subjective measures of awareness further showed that, while participants were aware of having acquired knowledge, they were at least partially unaware of what knowledge they had acquired. In other words, incidental exposure had resulted in the acquisition of both, conscious

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. 255–275). Honolulu: University of Hawai‘i, National Foreign Language Resource Center.

(explicit) and unconscious (implicit) knowledge. These results are consistent with previous studies using subjective measures to investigate the implicit and explicit learning of novel words (e.g., Hamrick & Rebuschat, 2012, in press) and L2 syntax (e.g., Rebuschat, 2008; Rebuschat & Williams, 2012). The results also demonstrate the benefit of employing subjective measures of awareness and of utilizing trained control groups.

Introduction

Implicit learning, essentially the process of acquiring unconscious (implicit) knowledge, is a fundamental aspect of human cognition (see Perruchet, 2008; Reber, 1993; Shanks, 2005, for overviews). 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; Dienes, 2012; Reber, 1993). The term *implicit learning* was first employed by Arthur Reber (1967) to describe a process during which participants derive knowledge from a complex, rule-governed stimulus domain without intending to and without becoming aware of the knowledge they have acquired. The term *explicit learning* is usually applied to a mode of learning that usually occurs under conditions in which participants are instructed to actively look for patterns; that is, learning is intentional, a process which tends to result in conscious (explicit) knowledge (though see footnote 4).

The field of second language acquisition (SLA) has a long-standing interest in the topic of implicit and explicit learning (see DeKeyser, 2003, and Williams, 2009, for reviews). One of the central questions concerns the possibility of second language (L2) learning without awareness, and Richard Schmidt's seminal publications have had a profound impact, both in terms of theoretical debates (e.g., Robinson, 1995, 2003; Schmidt, 1990, 1993, 1994, 1995, 2001; Tomlin & Villa, 1994) and empirical inquiry (e.g., Faretta-Stutenberg & Morgan-Short, 2011; Hama & Leow, 2010; Leow, 1997, 1998, 2000; Leung & Williams, 2011a, 2011b; Rosa & Leow, 2004; Rosa & O'Neill, 1999; Sachs & Suh, 2007; Williams, 2004, 2005). Schmidt (1995, p. 29) distinguishes two levels of awareness. Awareness at the level of noticing refers to the "conscious registration of an event," whereas awareness at the level of understanding refers to the "recognition of a general principle, rule or pattern." Registering instances of the morpheme—*ed* (*walked*, *jumped*, *played*, etc.) in a text would be an example of noticing. Recognizing that—*ed* indicates the past tense or, further, recognizing that there is a very productive rule underlying these instances (*add an—ed to produce the regular past tense*) would be awareness at the level of understanding. According to Schmidt (1990, 1993, 1994, 1995, 2001), only awareness at the level of noticing is required for initial processing of novel input: Noticing is "the necessary and sufficient condition for the conversion of input into intake" (Schmidt, 1993, p. 209).

Schmidt's proposals have been highly influential in SLA research, and the view that awareness is important (if not essential) for learning to take place has received empirical support from a series of studies (e.g., Leow, 1997, 1998, 2000; Rosa & Leow, 2004; Rosa & O'Neill, 1999). For example, Leow (2000) investigated the relationship between awareness (or the lack thereof) and L2 learners' subsequent recognition and written production of morphological forms (irregular 3rd person singular and plural preterit forms of stem-changing—*ir* verbs). L2 learners of Spanish completed an ingenious crossword-puzzle task that made the target feature available to participants. They were then tested to see whether learning took place. Importantly, participants were instructed to think aloud while completing the experimental tasks. Leow (2000, pp. 564–565) used these concurrent verbal reports to classify participants into aware and unaware groups: "Any participant who provided a report of being aware of the targeted forms or some form of metalinguistic description of the underlying rule would be assigned to the aware group; participants failing to fulfill the criteria would be assigned to the unaware group." The results indicated that only learners who were aware of the target forms improved from pretest to posttest; learners who were unaware of the target forms did not improve at all. Based on these findings, Leow (2000) suggested that awareness plays a crucial role in L2 acquisition by making input available for subsequent processing. Leow's observation has received support from a series of studies using the same think-aloud methodology (Leow, 1997, 1998; Rosa & Leow, 2004; Rosa & O'Neill, 1999). It seems well established that higher levels of awareness are generally associated with greater demonstrations of learning (Hamrick & Rebuschat, 2012; Rebuschat & Williams, 2012; Rosa & Leow, 2004; Rosa & O'Neill, 1999; Sachs & Suh, 2007).

While it is generally accepted that attention and awareness play an important role in learning (see Leow & Bowles, 2005; Robinson, 2003; Schmidt, 2001, for reviews), the assumption that low levels of awareness of linguistic phenomena are necessary for their acquisition has been challenged in recent years. In a widely cited study, Williams (2005) examined the acquisition of an artificial determiner system in a meaning-oriented task. Participants were exposed to four new determiners (*gi*, *ro*, *ul*, and *ne*) which encoded both distance (near vs. far) and animacy (animate vs. inanimate). At the beginning, participants were told that the determiners functioned like English determiners, except that they also encoded distance, for example *gi* and *ro* were used for near objects, while *ul* and *ne* were used for far objects.[1] Participants were not informed that the artificial determiners also encoded animacy: *gi* and *ul* were used with animate objects, whereas *ro* and *ne* were used with inanimate ones. The role of animacy in determiner usage thus served as a hidden regularity.

Participants were exposed to the semi-artificial language under incidental learning conditions; that is, they did not know they were going to be tested. In the training phase, participants were instructed to listen to each training sentence (e.g., "I spent an hour polishing *ro* table before the dinner party") to indicate whether the novel determiner used in the sentence meant *near* or *far*, to repeat the sentence verbatim, and to form a mental image of the general situation described by the picture. The testing phase consisted of two parts. In the first part, participants read part of a novel sentence such as "The lady spent many hours sewing..." and then had to select the appropriate segment to complete it from two options which matched in their distance values and differed only according to animacy, e.g., "... *gi* cushions / *ro* cushions." Participants were then interviewed to gauge their awareness of the animacy regularity. In the second part, those participants who were still unaware of the relevance of the animacy feature were given the same test sentences but this time with the instruction to discover the rules that determined the choice of determiners. They were then interviewed again to assess the conscious or unconscious status of any acquired knowledge.

Williams (2005) found that, after the first part of the test, 80% of participants reported to be unaware of the relevance of animacy in determiner usage, despite performing at 61% accuracy (significantly above chance) in the sentence completion task. After the rule discovery task, 50% of participants were still unaware of the rule, yet their accuracy was still significantly above chance (58%). The results were interpreted as demonstrating that participants can acquire form-meaning connections without becoming aware of what those connections are. In other words, learning without awareness was taken to be possible.

1 In Williams (2005), half the participants were told that *gi* and *ro* were used for near objects and *ro* and *ul* for distant ones, while the other half was told the opposite.

Williams (2005) was recently the target of an important extension study. Hama and Leow (2010) adapted the methodology of Williams (2005) to assess whether learning without awareness is, in fact, possible. According to Hama and Leow, the discrepancy between Leow (2000) and Williams (2005) can be explained by methodological differences. While the former study employed think-aloud protocols to assess awareness during the training and testing phases (in addition to probe questions after each phase, e.g., "Did you notice anything interesting about the verbs?"), the latter study relied on retrospective verbal reports. As a consequence, Leow (2000) investigated the role of awareness at the time of encoding, whereas Williams (2005) examined whether exposure had resulted in conscious or unconscious knowledge. In other words, Leow (2000) focused on the process of learning while Williams (2005) focused on the product. Hama and Leow's (2010) replication of Williams (2005) modified the original design by adding think-aloud protocols to the experimental tasks, i.e., participants were prompted to verbalize their thoughts while performing the tasks. They also included a production task and changed the forced-choice test to include four options instead of two. Finally, they also kept all of the tasks in the auditory modality, in contrast to Williams (2005), who had used the auditory modality for training and the written modality for testing.

Forty-three native speakers of English were exposed to the artificial determiner system employed by Williams (2005) by means of the same exposure task. Afterwards, participants were asked to perform the two tests (multiple choice recognition, production). The recorded verbal reports were transcribed and coded as *understanding*, *noticing*, or *no report* (see Rosa & O'Neill, 1999). A verbal report was coded as *noticing* when some aspect of animacy was mentioned or commented upon, *understanding* when correct rules related to animacy were mentioned, or *no report* when the report did not fall under the coding categories of noticing or understanding. Hama and Leow (2010) found no evidence for awareness of animacy during the training phase. However, the think-aloud protocols for the test phase clearly provided evidence for awareness at the level of noticing and at the level of understanding. Based on the data, nine participants were classified as aware of the hidden regularity and 34 as unaware. Further analyses indicated a significant learning effect in the aware group on both tests but no learning effect in the unaware group; that is, learning was restricted to those participants who became aware of the hidden regularity. Hama and Leow (2010) concluded that there was no evidence for learning without awareness. These results are supported by another extension study, Faretta-Stutenberg and Morgan-Short (2011).

Despite measuring awareness at different stages of the learning process, the methodologies employed by Williams (2005) and Leow (2000; Hama & Leow, 2010) share a basic limitation: They both rely on verbalization (or lack thereof) to disentangle implicit and explicit processes (Leow) and knowledge (Williams). In the case of concurrent reports (think-aloud protocols), it is assumed that learning proceeds without awareness if participants do not verbalize relevant features of the target system while engaged in either the training task or the test task. As Schmidt (2001, p. 20) writes, "the clearest evidence that something has exceeded the subjective threshold and been consciously perceived or noticed is concurrent verbal report."[2] In the case of retrospective reports, it is assumed that knowledge is unconscious when participants show an effect of training (e.g., above-chance performance on a grammaticality judgment task), despite being unable to describe the knowledge that

2 Schmidt (2001) is not arguing here that verbalization is an exhaustive index of awareness. He is merely highlighting that, when participants do verbalize knowledge, one can be confident that this knowledge is consciously represented. Studies of awareness and learning, however, sometimes appear to assume the contrapositive (i.e., if something is not verbalized, then it has not been noticed), which does not follow from Schmidt's (2001) observation.

underlies their performance. In both cases these assumptions are probably not warranted. For example, awareness may happen more quickly than concurrent verbalization allows expression of, given that "subjective awareness is fleeting and cannot be completely recorded" (Schmidt, 1995, p. 28). In retrospective verbalization, awareness may have decayed in memory by the time participants are asked to report on it. In addition, participants might fail to report knowledge simply because they lack confidence or do not realize that the knowledge is relevant. When participants are given the option of not responding during retrospective or concurrent verbal reports, then conscious knowledge, though present, may simply not be detected. Erdelyi and Becker (1974, cited in Dienes & Berry, 1997) also report that participants who are unable to verbalize knowledge on their first attempt are often able to do so when prompted again at a later point in time. Verbal reports might thus not be sensitive enough to capture all of the relevant conscious knowledge.

While lack of verbalization does not provide strong evidence for learning without awareness (in the case of think-alouds) or implicit knowledge (in the case of retrospective reports), it is important to note that the presence of verbalization does not automatically mean that all learning in the experiment involved awareness or that only explicit knowledge was acquired. Both procedures lack exclusivity in the sense that they might be contaminated by unconscious knowledge (Reingold & Merikle, 1990). When think-aloud data indicates that participants were aware of a given complex L2 phenomenon, this does not necessarily mean that other aspects of the same phenomenon have not been acquired without awareness. In addition, one needs to ask what processes contributed to participants suddenly becoming aware of a feature in the first place, with implicit processing (e.g., in the form of associative or statistical learning) a possible candidate in this case. Likewise, when participants verbalize knowledge at the end of the experiment, this does not mean that participants only acquired conscious knowledge. In fact, recent research (Hamrick & Rebuschat, 2012, in press; Rebuschat, 2008; Rebuschat & Williams, 2012, 2013; Tagarelli, Borges Mota, & Rebuschat, 2011) suggests that, even under incidental learning conditions, participants are quite likely to acquire both implicit and explicit knowledge.[3]

The present study is the third extension of Williams (2005), after Hama and Leow (2010) and Faretta-Stutenberg and Morgan-Short (2011). Our primary objective is to contribute to the debate on awareness and language learning by illustrating the usefulness of a measure of awareness that does not rely on verbalization. Before discussing the details of our experiment, we briefly introduce the subjective measures of awareness we employed in order to assess the conscious status of learners' knowledge in our study. For a more comprehensive review, see Rebuschat (in press).

Subjective measures of awareness

Dienes (2004, 2012) 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 participants to perform 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,

3 Using subjective measures of awareness, these studies indicate that participants acquire both conscious (explicit) and unconscious (implicit) knowledge as a result of exposure and that the type of knowledge seems to depend on the learning context. Under incidental conditions, participants develop primarily unconscious knowledge, while under intentional conditions, they develop primarily conscious knowledge.

memory, rule knowledge). Knowledge can be considered unconscious if participants believe they are guessing when their classification performance is, in fact, significantly above chance. Dienes, Altmann, Kwan, and Goode (1995) called this the *guessing criterion*. Knowledge can also be considered unconscious if participants' confidence is unrelated to their accuracy. This criterion, introduced by Chan (1992), was labeled the *zero correlation criterion* by Dienes et al. Several studies have shown that performance on standard AGL tasks can result in unconscious knowledge according to these criteria (e.g., Dienes et al., 1995).

Dienes (2004) suggested that, when participants 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, participants use their structural knowledge to construct a different type of knowledge, namely whether the test items share 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 conscious only if it is explicitly represented, that is if there is a higher-order thought such as *I know/think/believe that an R can be repeated several times* (Rosenthal, 2005). Likewise, judgment knowledge is conscious only 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 judgments without knowing why. Here, structural (linguistic) knowledge is unconscious while 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. If, on the other hand, both structural and judgment knowledge are unconscious, the phenomenology is that of guessing. In both cases (guessing and intuition), the structural knowledge acquired during training is unconscious. The experiment below exemplifies how subjective measures can be employed to investigate whether incidental exposure to a new determiner system can result in unconscious knowledge in adult learners.

Methods

The present study had two objectives. The first objective was to confirm that adult learners can acquire novel form-meaning mappings incidentally as a result of exposure in a meaning-oriented task and that they are able to generalize this knowledge to novel stimuli. The second, more important objective was to illustrate the usefulness of subjective measures of awareness and to compare the latter to the information gained via retrospective reports.

Participants

Thirty undergraduate students at Georgetown University participated in the study (18 women, 12 men; Mage=20) and were either assigned to the experimental group (*n*=15) or the trained control group (*n*=15). A *trained* control condition was used following recent calls for more robust procedures to ascertain learning (e.g., Hamrick, 2012, 2013; Reber & Perruchet, 2003). Trained control groups receive training conditions that are identical to experimental groups but with the relevant independent variables randomized and balanced, rather than removed altogether. The logic behind this procedure stems from the notion that all participants have unforeseen response biases in test phases based on their prior knowledge (Reber & Perruchet, 2003). These biases are "noise" that influences test performance beyond what is learned during training. The use of trained controls ensures that such noise can be identified and accounted for, allowing the effects of the independent variable(s) to be isolated.

All participants were native speakers of English. Twelve participants reported having an additional native language; these included Farsi, French, German, Korean, Mandarin, and Spanish. Twenty-seven participants had studied a foreign language, including Spanish (18), French (12), German (7), Arabic (4), Korean (4), Latin (4), Italian (3), Mandarin (3), Russian (2), Catalan (1), Japanese (1), and Portuguese (1). Only three participants said they did not know any foreign languages. Experimental and control groups did not differ significantly with regard to age, gender, number of linguistics courses taken, number of foreign languages studied, or university year, all $p > .05$. Eleven participants were pursuing degrees in linguistics (seven in the experimental group and four in the control group). Participants who were enrolled in a linguistics course were offered 5% extra credit on a homework assignment in that course for their participation.

Materials

The artificial determiner system used in this experiment was taken from Williams (2004, 2005) (see also Faretta-Stutenberg & Morgan-Short, 2011; Hama & Leow, 2010). The system consists of four artificial determiners (*gi*, *ro*, *ul,* and *ne*) which encode both distance (near vs. far) and animacy (animate vs. inanimate). The determiners *gi* and *ro* precede nouns that refer to objects that are near, while *ul* and *ne* are used for nouns that refer to distant objects. In addition, *gi* and *ul* are employed to refer to animate entities (natural, living, moving things), whereas *ro* and *ne* are used with inanimate ones (man-made, non-living, stationary things). As in the previous studies, participants were trained explicitly on the near/far distinction but were not told of the regularity involving animacy. The training and test sentences employed in the current experiment are available in the IRIS digital repository (www.iris-database.org). The form-meaning mappings are illustrated in Table 1.

Table 1. Artificial determiner system used in the present study

	near	**far**
animate	*gi*	*ul*
inanimate	*ro*	*ne*

Training items

The noun phrases (NPs) used in the exposure phase for the experimental group can be found in Table 2. As in Williams (2005, Experiment 2), the training set included 12 animate and 12 inanimate nouns, each of which was presented in both its singular and its plural form. There were thus 48 items in total (24 singular, 24 plural). Each noun, in both singular and plural forms, only appeared with one determiner (e.g, *gi bear* and *gi bears*). This was done to ensure that participants who demonstrated improvement would be doing so on the basis of their learning of form-meaning connections (e.g., that *gi* is used with animate nouns) as opposed to, perhaps, form-form associations between determiners (e.g., that any noun that takes *gi* [near] also takes *ul* [far]).

Table 2. Noun phrases (48) employed in the exposure phase of the experimental group

animate		inanimate	
near	**far**	**near**	**far**
gi bear(s)	*ul* bee(s)	*ro* box(es)	*ne* book(s)
gi cow(s)	*ul* bird(s)	*ro* cup(s)	*ne* clock(s)
gi dog(s)	*ul* cat(s)	*ro* picture(s)	*ne* cushion(s)
gi lion(s)	*ul* fly (flies)	*ro* sofa(s)	*ne* plate(s)
gi pig(s)	*ul* monkey(s)	*ro* table(s)	*ne* stool(s)
gi rat(s)	*ul* snake(s)	*ro* television(s)	*ne* vase(s)

note. The nouns appeared in both singular and plural forms.

The training set was subdivided into two sets, with six NPs of each type (near/animate, near/inanimate, far/animate, far/inanimate) per set. The same determiner-noun combinations were used in Set 1 and Set 2, but they differed in terms of grammatical number. That is, if a given noun appeared in its singular form in Set 1, then the same noun appeared in the plural in Set 2, and vice versa. Each set (and therefore each sentence) was repeated three times during training so that subjects were exposed to a total of 144 items, the same number as in Williams (2005). The sets were presented in alternating order (Set 1, Set 2, Set 1, Set 2, Set 1, Set 2), and the ordering of the sentences within each set was randomized on each iteration. Some new sentences also had to be created so that nouns could be trained in both their singular and plural forms.

In order to allow for the inclusion of a trained control group, an additional set of training materials was developed. The 48 sentences for the trained controls were identical to those of the experimental participants except that the animacy regularity was removed by changing which determiners were used with which nouns in such a way that none of the determiners was reliably associated with a particular animacy value. That is, each determiner was used half of the time with animate nouns and half the time with inanimate nouns. Since all participants were pre-trained explicitly on distance, the near/far meanings of the determiners were maintained. Care was also taken to ensure that each determiner appeared half the time with a singular noun and half the time with a plural noun. Table 3 contains the items used during the exposure phase for the trained control group.

Table 3. Noun phrases (48) employed in the exposure phase of the trained controls

animate		inanimate	
near	**far**	**near**	**far**
gi bear *gi* bears	*ul* bee *ne* bees	*gi* pictures *ro* picture	*ul* cushions *ne* cushion
ro rat *ro* rats	*ul* bird *ne* birds	*gi* sofa *ro* sofas	*ul* clocks *ne* clock
gi dog *gi* dogs	*ul* cat *ne* cats	*gi* boxes *ro* box	*ul* stool *ne* stools

ro pig	*ul* flies	*gi* cups	*ul* vases
ro pigs	*ne* fly	*ro* cup	*ne* vase
gi lion	*ul* monkeys	*ro* tables	*ul* books
gi lions	*ne* monkey	*ro* table	*ne* book
ro cow	*ul* snake	*gi* television	*ul* plate
ro cows	*ne* snakes	*gi* televisions	*ne* plates

Test items

The testing set, which was the same for both groups, contained completely new sentence contexts, none of which had appeared during the training. The sentences were designed to test three types of NPs: *trained*, *partially trained*, and *new*. For the experimental group, the trained NPs had already occurred in exactly the same form in the exposure phase (e.g., *gi bears*). In the case of the partially-trained NPs, the determiner and the noun had occurred during training but not in this specific configuration. For example, if *ro picture* (the near picture) had occurred in training, then either the singular or plural version of the noun *picture* was presented in a far context, requiring *ne*, on the test. Finally, the new NPs were items in which the noun had not occurred at all during the exposure phase (e.g., *gi rabbit*). It is important to note that Williams (2004, 2005), Hama and Leow (2010), and Faretta-Stutenberg and Morgan-Short (2011) only featured two types of test items, namely trained and partially trained (which they call "generalization"). In contrast to the present experiment, their studies did not contain true generalization items.

Table 4. Noun phrases employed in the testing phase

	animate		**inanimate**	
	near	**far**	**near**	**far**
trained	*gi* rats	*ul* bees	*ro* cups	*ne* cushions
trained	*gi* cow	*ul* flies	*ro* television	*ne* clocks
trained	*gi* dog	*ul* cat	*ro* sofa	*ne* book
partially trained	*gi* monkeys	*ul* bears	*ro* plates	*ne* pictures
partially trained	*gi* snakes	*ul* lion	*ro* stools	*ne* table
partially trained	*gi* bird	*ul* pig	*ro* vase	*ne* box
new	*gi* elephants	*ul* camels	*ro* desks	*ne* candles
new	*gi* hamster	*ul* horses	*ro* spoons	*ne* lamp
new	*gi* rabbit	*ul* turtle	*ro* phone	*ne* towel

note. Trained items were noun phrases repeated from training. Partially-trained items feature nouns that occurred during training but with a different determiner. New items feature nouns that did not occur during training.

There were 36 test items—novel sentences taken mostly from Williams (2005) and Hama and Leow (2010)—with six of each type (*trained*, *partially trained*, and *new*) for each animacy class.[4] Plurality and distance values were balanced within each test-item type so that, for example, the six trained animate items included three singular and three plural NPs as well as three near and three far NPs, taking care not to confound plurality and distance.

4 Some of the sentences from Williams (2005) were modified to follow North American English as opposed to British English.

The trained and partially-trained items were the same as the ones used by Williams (2005), with the exception of the noun *rat*, which we included instead of *mouse* in order to avoid irregular plurals. Most of the new NPs in the current study contain nouns taken from Hama and Leow (2010), plus four new nouns (*hamster*, *camel*, *towel*, *desk)*, which we added for counterbalancing purposes (and to test generalization ability). Table 4 displays the NPs used in the testing phase.

Procedure

The experiment consisted of (i) informed consent, (ii) vocabulary pre-training, (iii) a training phase, (iv) a testing phase, and (v) debriefing. Participants met individually with one of the researchers in a quiet laboratory setting. The training and assessment tasks were run on Apple iMac computers and delivered via Cedrus SuperLab Pro (version 4.0.7b). Participants were also audio-recorded while performing the tasks (by means of a handheld digital recorder placed on the table) to ensure that they had followed the instructions. The experiment concluded with a debriefing session during which participants completed a post-exposure interview (retrospective verbal reports) with one of the authors and a background questionnaire. The entire session took approximately one hour.

Vocabulary pre-training

As in Williams (2005) and Hama and Leow (2010), a vocabulary pre-training activity introduced participants to the four novel determiners and their English translations. The activity was administered via Microsoft PowerPoint. Participants were told that they were going to learn four new words (*gi, ro, ne, ul*) and were presented with a list of the words and their respective distance values in English (*gi* and *ro*=near, *ne* and *ul*=far). Participants then completed two practice tasks that exposed them to 12 written repetitions of each novel word. In the first task, participants saw the four determiners and their English translations on the screen, but question marks appeared for one of the determiners (e.g., *ne*=far, *??*=far, *gi*=near, *ro*=near). They were instructed to say aloud the missing determiner (in the example, *ul*). Afterwards, they pressed a mouse button and the missing determiner would appear; that is, they received feedback. In the second task, participants were presented with one of the artificial determiners on the screen. Their task this time was to say aloud the English translation of the novel word. After saying the word, participants clicked on the screen, and the correct translation would appear. Presentation order was randomized and repeated five times for each determiner. The remaining four exposures occurred during the instructions and examples provided for each portion of the pre-training exercise.

Participants were encouraged to complete the vocabulary pre-training at their own pace and more than once if they desired and were informed that the pre-training would be immediately followed by a short test to evaluate whether they had successfully learned the four novel words. One participant in the control group and two in the experimental group chose to repeat the pre-training once. The quiz was administered using the online survey and testing website *ClassMarker* (www.classmarker.com). Participants were required to score 90% or higher on the quiz in order to move on to the training phase, and all were able to do this on their first attempt.

Training phase

Experimental participants and trained controls were provided with written instructions, explaining the general purpose of the experiment. These were based on the instructions provided by Williams (2005, pp. 281–282) with only minor changes. Participants were informed that the four artificial determiners functioned like the English word *the*, except that they also encoded distance: *gi* and *ro* were used for near objects, while *ul* and *ne* were used for far objects. Importantly, participants were not informed that *gi* and *ul* were used with animate objects, whereas *ro* and *ne* were used with inanimate ones. A sample sentence (*The little boy patted gi tiger in the zoo*), which did not recur during the training task, illustrated how the determiners could be used in a sentence context.

Participants were then told that they would be presented with written sentences that included the new words they had learned during the pre-training phase and that their task was to read the sentence aloud and then indicate, as quickly and accurately as possible, whether the novel word meant *near* or *far* by pressing the corresponding key (marked with a sticker) on the computer's keyboard. (Note that, unlike Williams, 2005, and Hama and Leow, 2010, the training sentences were presented visually on the screen, not auditorily.) After each decision they were also asked to repeat the novel word together with its noun (*gi tiger*, in the example above), while simultaneously forming a mental image of the situation. For example, in the sample sentence, participants were encouraged to imagine a boy patting a tiger that was close to him. Forming a mental image of the sentence context, including the relationship between the novel word and the following noun, encouraged participants to process the meanings of the words and the overall situation described in the sentence. This was emphasized as an important aspect of the experiment by the researcher administering the treatment. Trained control participants received the same instructions and procedure. The only difference was that the determiners they saw were randomized and counterbalanced so that no determiner ever reliably indicated animacy. All participants completed a short practice session with four sentences that were not repeated in the training phase.

Testing phase

After the exposure phase, participants were visually presented with 36 completely new sentences. For each test item, the computer displayed a sentence context (e.g., *The boy patted ___ tiger in the zoo*) with two choices of artificial determiners (e.g., *gi* and *ro)* located in the bottom left and right corners of the screen. Importantly, these options always matched in their distance values (*gi* and *ro* can both refer to near entities) while differing in their animacy values, which means that participants could no longer respond only according to the distance information they had been explicitly taught during the vocabulary pre-training phase (and which they had been instructed to focus on while processing the sentences of the training phase). The participants were instructed to read through the entire sentence and to "choose the word that seems more familiar, better, or more appropriate," based on what they had done so far. They could enter their choice by pressing a corresponding key on the keyboard. As in Williams (2005), there were two response options. (Hama and Leow, 2010, provided participants with four options.)

After selecting which word best filled the blank, participants were also asked to indicate how confident they were in their decision and what the basis of their decision was. These confidence ratings and source attributions, respectively, served as subjective measures of awareness. Participants could indicate their levels of confidence by selecting one of four response options for each item: *complete guess*, *somewhat confident*, *very confident*, or *absolute certainty*. Participants were instructed to select the *complete guess* category only if they had no confidence whatsoever in their classification decision and truly believed to be guessing. If they had some degree of confidence, they were asked to select the *somewhat confident* or the *very confident* categories instead. If they were 100% certain that their classification was correct, then they were instructed to select the *absolute certainty* category. In the case of the source attributions, participants were asked to select one of four response options: *guess*, *intuition*, *memory*, or *rule knowledge*. Participants were instructed to use the *guess* category only when decisions were based on real guesses; that is, they might as well have flipped a coin. The *intuition* category was to be selected if participants had a gut feeling that they were right but did not know why. The *memory* category was to be selected when judgments were based on the recollection of items from the earlier part of the experiment. Finally, the *rule knowledge* category was to be selected following decisions that were based on a rule that participants would be able to report at the end of the experiment. All participants were provided with these instructions before starting the testing phase. (Note that we had only one test phase, in contrast to Williams, 2005, who featured two.) Participants completed a short practice session with four sentences that were not repeated in the test phase.

The 36 test sentences were presented in the same order for all participants. As in Williams (2005), they were arranged so that participants would not be able to make animacy comparisons across adjacent items with the same distance values. For example, test items targeting far animate NPs (e.g., *ul bees*) were never followed by far inanimate NPs (e.g., *ne clocks*). Because our study featured three types of test items (as opposed to two in the two previous studies), we could not use exactly the same item ordering as that employed by Williams (2005). However, we did follow his ordering on the more abstract level of plurality, distance, and animacy features.

Debriefing session

Following the testing phase, all participants completed a short interview with one of the researchers. Participants were first asked what criteria they had used to make their choices. If they made any references to living/nonliving, moves/doesn't-move, or a similar distinction, they were asked at what point they had become aware of this difference. The participants were then asked whether they had ever indicated rule knowledge as a basis for their decisions. If so, they were asked to describe what they had been thinking and why they had selected the *rule knowledge* category. If they had not indicated rule knowledge as a source, they were prompted to share any other ways in which they had made their choices, whether on the basis of intuition or other sources. If, up to this point in the interview, participants had not mentioned anything related to animacy or had not reported indicating rule knowledge as a basis for their decisions, they were informed that there was a rule and were asked to speculate about what the rule might have been. If animacy was still not mentioned, the researcher explained the system, and then asked participants if they had considered the possible relevance of animacy at any point during the training or assessment task.

In addition to the retrospective verbal reports, participants also completed a brief questionnaire asking for their age, field of study, previous experience in linguistics courses, native language(s), and foreign languages studied. Where applicable, participants provided additional information regarding their foreign language background, including contexts of instruction, levels of formal schooling, length of study, and self-reported proficiency.

Results

As in Williams (2005), performance on the two-alternative forced-choice (2AFC) task served as the measure of learning. The confidence ratings, source attributions, and retrospective verbal reports were used to determine to what extent participants were aware of having acquired knowledge and whether the acquired knowledge was conscious or not. It is worth remembering that Williams (2005) only relied on retrospective verbal reports.

Performance on the 2AFC task

Overall performance of experimental and control groups

The analysis of the test data showed that the experimental group identified 74.8% (SD=28.3%) of the test items correctly and the trained control group 49.3% (SD=9.6%). Levene's statistic showed that variances were not homogeneous. The adjusted independent sample t-test indicated that the experimental participants significantly outperformed the trained controls, $t(17.16)=3.31$, $p<.05$. Further analysis showed that the experimental group performed significantly above chance on this task, $t(14)=3.39$, $p<.05$, while the controls scored at chance, $p>.05$. That is, the training phase produced a clear overall learning effect in the experimental participants but not a clear overall learning effect in the trained controls.

As described above, eleven participants (seven experimental, four control) were pursuing a degree in linguistics (major or minor). Given the potential impact of this background on performance in the experiment (see Williams, 2004, 2005), we decided to compare these students to those who were not studying for a degree in linguistics. The analysis showed that, in the experimental group, the non-linguistics students identified 79.3% (SD=18%) of the test items correctly and linguistics students 69.4% (SD=37.8%). In the control group, the non-linguistics students performed at 51% (SD=10.2%) and the linguistics students at 46% (SD=8.4%). Our analysis showed that there were no significant differences in performance between linguistics participants and non-linguistics participants, either in the experimental group, $t(13)=.66$, $p>.05$, or the control group, $t(13)=.87$, $p>.05$.

Performance of experimental and control groups across different test-item types

The experimental group correctly identified 78.3% (SD=9.5%) of the trained NPs, 73.3% (SD=7.5%) of the partially-trained NPs, and 72.8% (SD=10%) of the new NPs. The trained controls correctly identified only 41.1% (SD=17.3%) of the trained NPs, 52.2% (SD=13.3%) of the partially-trained NPs, and 54.4% (SD=5.6%) of the new NPs. The experimental group performed significantly above chance on trained NPs, which had already occurred in the exposure phase (though in different sentence contexts), $t(11)=10.35$, $p<.001$, as well as on partially-trained NPs, whose noun had occurred with a different determiner during training, $t(11)=10.74$, $p<.001$, and also on new (generalization) NPs, which had not occurred during training at all, $t(11)=7.86$, $p<.001$. The trained controls performed significantly above chance on new NPs, $t(11)=2.76$, $p<.05$, but not on trained or partially-trained NPs, $p>.05$.

A 2x3 mixed ANOVA with Group (2 levels: Experimental, Control) as the between-subjects variable and Test-Item Type (3 levels: Trained, Partially Trained, New) as the within-subjects variable revealed no effect of Test-Item Type, but there was a significant main effect of Group, $F(1, 22)=97.89$, $p<.001$, $\eta p2=.82$, and a significant Group*Test-Item Type interaction, $F(2, 44)=4.87$, $p<.05$, $\eta p2=.18$. To follow up on the significant interaction effect, we performed additional ANOVAs comparing the groups (Experimental, Control) on each type of test item. These revealed that the experimental group was significantly more accurate than the control group on all three test-item types: trained, $F(1, 23)=42.90$, $p<.001$, partially trained, $F(1, 23)=22.95$, $p<.001$, and new (corrected with Welch's F), $FW(1, 17.18)=17.18$, $p<.001$. We also performed repeated-measures ANOVAs to establish if there were differences within each group across the test-item types (Trained, Partially Trained, New). In the experimental group, there was no significant effect of Test-Item Type, $F(2, 22)=1.40$, $p>.05$, which shows that, in this group, participants performed similarly across the types of test items. However, in the control group, performance did differ significantly across test-item types, $F(2, 22)=3.48$, $p<.05$, $\eta p2=.24$. Contrasts showed there was a significant difference between trained and new items, $F(1, 11)=8.25$, $p<.05$, $\eta p2=.43$, but no significant difference between trained and partially trained items, $F(1, 11)=2.57$, $p>.05$, or between partially trained and new items, $F(1, 11)=.27$, $p>.05$. Figure 1 illustrates the performance of the two groups across the three types of test items.

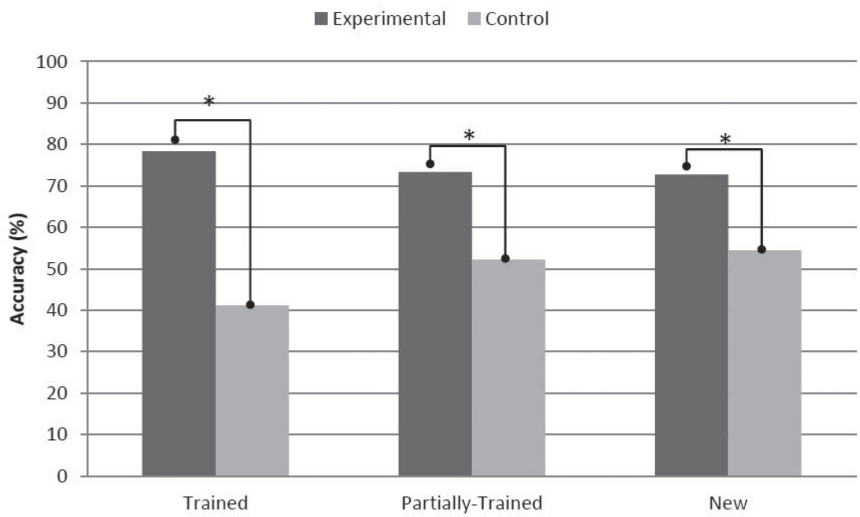

Figure 1. Test performance of experimental and control participants on trained, partially-trained, and new items

Retrospective verbal reports

The following analyses focus on the experimental group, given that there was no overall learning effect in the trained controls. The verbal report data from two experimental participants was not available for analysis due to technical failure. The analysis of the remaining data showed that nine participants were able to verbalize at least some knowledge regarding the animacy regularity. In the analyses below, these participants will be referred to as the *aware group*. The four remaining experimental participants expressed no awareness of the animacy regularity. These will be referred to as the *unaware group*. Below we report the performance of both groups on the 2AFC task.

Overall performance of aware and unaware experimental participants on the 2AFC task
The aware group identified 79.6% (SD=31.6%) of the test items correctly and the unaware group 53.5% (SD=10.7%). Aware participants performed significantly above chance, $t(8)=2.82$, $p<.05$, while unaware participants were indistinguishable from chance, $p>.05$. A t-test showed that aware participants did not significantly outperform unaware participants, $p>.05$. However, this result is likely an uninteresting effect of the small number of participants in the unaware group (n=4). The analysis of the retrospective verbal reports suggests that learning in this experiment was restricted to those participants who were able to verbalize at least some knowledge related to the hidden animacy regularity. In contrast to Williams (2005), the verbal reports thus provide no evidence of unconscious knowledge in the experimental group (though see below).

Performance of aware and unaware participants across different test items
The aware group identified correctly 82.4% (SD=11%) of the trained items, 79.6% (SD=6.4%) of the partially-trained items, and 76.9% (SD=11%) of the new items. In all cases, these

scores were significantly above chance, *ps*<.001. The unaware group identified correctly 64.6% (*SD*=41.9%) of the trained items, 54.2% (*SD*=33.4%) of the partially-trained items, and 56.3% (*SD*=35.6%) of the new items. However, none of the scores were significantly above chance, *ps*>.05. Further analysis indicated that aware participants significantly outperformed unaware participants on partially-trained items, $t(11.81)=2.59$, $p<.05$, but there were no significant differences between groups on trained items or new items. Again, this is probably due to the small sample size of the unaware group, coupled with large standard deviations. Figure 2 illustrates the performance of the aware and unaware groups across the three types of items.

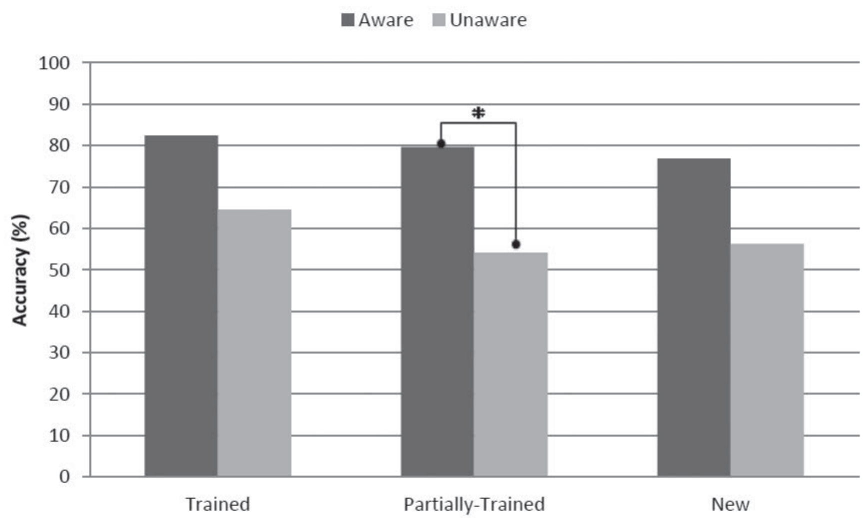

Figure 2. Test performance of aware and unaware participants on trained, partially-trained, and new items

Subjective measures

The following analyses of the subjective measures focus on the experimental group, given that we did not find an overall learning effect in the trained controls.

Confidence ratings

As shown in Table 5, in terms of proportion, experimental participants tended to select the *absolute certainty* category most frequently and the *complete guess* category 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 less accurate when reporting to be absolutely certain or somewhat confident. Accuracy was lowest for those grammaticality decisions in which participants had no confidence whatsoever and reported to be truly guessing. Experimental participants scored significantly above chance when reporting to be somewhat confident, very confident, and absolutely certain, *ps*<.05. However, when participants reported to be guessing, performance was indistinguishable from chance, $p>.05$, indicating that the guessing criterion for unconscious judgment knowledge was not satisfied. The fact that participants were more accurate when reporting higher levels of confidence (82.6% and 74.4%, respectively) also suggests the existence of conscious judgment knowledge. In other words, the confidence ratings suggest that participants in the experimental group were aware of having acquired some knowledge during the training phase. Table 5 summarizes the findings.

Table 5. Accuracy and proportion of responses (%) across confidence ratings

	complete guess	somewhat confident	very confident	absolute certainty
accuracy	50	72.6*	82.6*	74.4*
proportion	6	23	34	36

note. Significance from chance: * $p<.001$.

Source attributions

As shown in Table 6, in terms of proportion, experimental participants most frequently believed their classification decisions to be based on rule knowledge, followed by memory and intuition. The *guess* category was selected least frequently. In terms of accuracy, experimental participants scored highest when reporting to have relied on memory, followed by the *intuition* and *rule knowledge* categories. Participants were least accurate when basing decisions on guesses. Further analysis showed that participants performed significantly above chance across the four categories, *ps*<.05. The fact that participants performed significantly above chance when basing their decisions on guessing or intuition suggests that at least some of the acquired structural knowledge was unconscious. Table 6 summarizes the findings.

Table 6. Accuracy and proportion of responses (%) across source attributions

	guess	intuition	memory	rule
accuracy	66.7*	75.4**	84.1**	74.4**
proportion	7	24	21	48

note. Significance from chance: *$p<.05$, **$p<.001$

Discussion

The results of the experiment confirm that adult learners are able to establish novel form-meaning connections under incidental learning conditions without the benefit of feedback and after a relatively brief exposure period. As described above, the present study employed three types of test items (*trained* NPs, *partially-trained* NPs, and *new* NPs), in contrast to Williams (2005), Hama and Leow (2010), and Faretta-Stutenberg and Morgan-Short (2011), who did not include true generalization items. Experimental participants performed significantly above chance across all test-item types, including new items, indicating that they were able to generalize their knowledge and that exposure to the artificial determiner system resulted in a knowledge base that is, at least partially, abstract.[5]

The analysis of the retrospective verbal reports showed that 70% of participants were able to verbalize some knowledge regarding the hidden animacy regularity. Further analysis showed that only these aware participants were actually performing significantly above chance in the testing phase, with unaware participants (n=4) performing at chance. The evidence from the verbal reports thus indicates that learning in the experiment was restricted to those experimental participants who had acquired explicit knowledge. In other words, while

5 One could argue, of course, that exposure did not result in abstract knowledge but that abstraction occurred during the test phase, when participants compared stored exemplars to the test items. Grammaticality judgments would then be based on "abstract analogy" (Brooks & Vokey, 1991; Vokey & Higham, 2005).

learning in the experiments occurred under incidental conditions, participants developed explicit knowledge, some of which they could verbalize in the form of partial or complete metalinguistic rules. At the least, participants were aware of having acquired some form of knowledge, even if they had difficulty verbalizing it. These observations are consistent with Hama and Leow's (2010) findings.

Interestingly, the subjective measures of awareness (confidence ratings and source attributions) show that the picture is more complex than this. The analysis of the confidence ratings showed that experimental participants scored significantly above chance when reporting to be somewhat confident, very confident, or absolutely certain in their classification decision but only at chance when reporting no confidence whatsoever in the accuracy of a judgment. This suggests that the experimental group had acquired conscious judgment knowledge; that is, they had a sense of having acquired knowledge and of relying on this knowledge in the testing phase. The analysis of the source attributions showed that experimental participants performed significantly above chance across the four response categories, including responses based on guessing or intuition. When basing decisions on implicit categories (guess, intuition), participants were accurate in 71% of judgments. This finding indicates the presence of at least some unconscious structural knowledge, which is consistent with Williams (2005). Given that participants also performed significantly above chance when basing decisions on more explicit categories (memory or rule, with a combined accuracy of 80%), it seems that participants acquired both conscious and unconscious structural knowledge as a result of exposure. Taken together, the subjective measures thus indicate that, while participants were aware of having acquired knowledge, they were at least partially unaware of what knowledge they had acquired. These results are consistent with previous studies using subjective measures to investigate the implicit and explicit learning of novel words (e.g., Hamrick & Rebuschat, 2012, in press) and L2 syntax (e.g., Rebuschat, 2008; Rebuschat & Williams, 2012).

From a methodological perspective, the experiment confirms that sole reliance on retrospective verbal reports (as in Williams, 2004, 2005) is clearly insufficient for the study of implicit and explicit learning, as Williams (2005, 2009) clearly acknowledges. The analysis of our verbal reports showed that most participants were able to describe the rules of the artificial determiner system and that learning was apparently restricted to these aware participants. If we had not included more sensitive measures of awareness in the form of confidence ratings and source attributions, we might have erroneously concluded that exposure to the artificial system resulted exclusively in explicit knowledge. We propose that future explorations of implicit and explicit language learning could benefit from the inclusion of more sensitive ways of detecting implicit and explicit knowledge. Our research suggests that subjective measures could play a useful role, though it is clear that these are not the only solution and that subjective measures have their own methodological difficulties (see Rebuschat, in press, for a comprehensive discussion).

The experiment also extends previous work using trained control participants (Hamrick, 2012, 2013; Hamrick & Sachs, 2013). In the first SLA study to use this procedure, Hamrick (2012, 2013) incidentally exposed participants to semi-artificial sentences with either probabilistic syntactic structures (experimental condition) or randomized syntactic structures (trained control condition). The training sentences in the experimental condition contained transitional probabilities between categories of 67% and 33% (e.g., the probability that an NP was followed by a PP was 67%). In the control condition, the same training sentences were presented, but the syntax was randomized and all of the transitional probabilities were balanced at 25%. Thus, trained control sentences contained no probabilistic cues to syntactic structure. Interestingly, in the surprise grammaticality judgment task following exposure, trained controls outperformed experimental participants on one of the three syntactic target structures. Thus, the trained controls, who unlike the experimental group were not exposed to and therefore could not have learned the probabilistic structure, displayed a clear bias toward endorsing one of the syntactic structures. Such findings raise issues of what counts as a valid baseline for learning, especially considering other studies that show that controls may perform significantly below chance (e.g., Rebuschat & Williams, 2012). In the present study, the trained control group did not perform significantly above chance in terms of overall accuracy, though they did perform above chance on new items, which confirms again that the validity of 50% as a baseline should not simply be assumed.

Conclusion

The main objective of this study was to contribute to the ongoing debate on the implicit and explicit learning of languages by focusing on a central methodological issue, namely how to detect implicit and explicit knowledge (Ellis, 2005). Our extension of Williams (2005) confirmed that learners are able to establish novel form-meaning connections under incidental learning conditions (see also Faretta-Stutenberg & Morgan-Short, 2011; Hama & Leow, 2010; Leung & Williams, 2011a, 2011b). Importantly, our study also showed that incidental exposure can result in both implicit and explicit knowledge of language, which helps to shed light on the conflicting results obtained by Williams (2005) and Hama and Leow (2010). Williams (2005) might have underestimated the extent to which participants acquired explicit knowledge, given that his study relied on a relatively insensitive measure of awareness (retrospective verbal reports). On the other hand, Hama and Leow (2010) might have overestimated the role of explicit knowledge, given that they were unable to assess whether there was also implicit knowledge present in their aware participants.

There are outstanding issues that our study, due to the nature of its design, was unable to address. For example, it is uncertain *when* participants developed conscious knowledge. Participants could have become aware of the hidden regularity either during the training phase or during the test phase, when they were suddenly forced to make a choice between options differing only in their animacy values and when they were prompted with requests for source attributions that suggested that their responses might have been based on a rule. The think-aloud protocols of Hama and Leow (2010) suggest that their participants became aware during the test phase, as do the comments of several of our participants in their retrospective verbal reports. If this is the case, then it could well be that the acquired knowledge was implicit until it was required for the test phase, the nature of which could then have led participants to become aware of the target feature.

More research is also necessary to establish how the acquired knowledge is represented. Participants in our study are likely to have acquired abstract knowledge, given their performance on generalization items. However, it is not clear what the nature of this knowledge is (patterns, chunks, etc.). If participants had acquired the hidden animacy rule, their performance on those classifications attributed to (relevant) rule knowledge should have been significantly higher (close to 100%). Our results showed that they actually were slightly less accurate on judgments based on rule knowledge (74.4%) when compared to those based on intuition (75.4%) or memory (84.1%).[6] Finally, both of our measures of awareness assessed the conscious and unconscious status of the acquired knowledge; like Williams (2005), we focused

6 When reporting to be using rule knowledge, some participants may have relied on *micro-rules*, that is a partial but representative subset of the rules employed to generate the stimuli (Dulany, Carlson, & Dewey, 1984; Reber, 1993).

on the *product* of learning and not on the *process* of learning (see Hama & Leow, 2010). For this reason, we are unable to say much about the role of noticing in our study, though given that most experimental participants were able to verbalize some knowledge, it seems clear that noticing is likely to have played a role. The inclusion of a variety of awareness measures, ranging from off-line tasks such as retrospective verbal reports and subjective measures to online tasks such as think-aloud protocols, is likely to be necessary to move the discussion forward (see Rebuschat, Hamrick, Sachs, Riestenberg, and Ziegler, 2013).

References

Berry, D. C., & Dienes, Z. (1993). *Implicit learning: Theoretical and empirical issues*. Hove, UK: Lawrence Erlbaum.

Brooks, R. L., & Vokey, R. J. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 120*, 316–323.

Chan, C. (1992). *Implicit cognitive processes: Theoretical issues and applications in computer systems design* (Unpublished DPhil thesis). University of Oxford, UK.

DeKeyser, R. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 313–348). Oxford: Blackwell.

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. (2012). Conscious versus unconscious learning of structure. In P. Rebuschat & J. N. Williams (Eds.) *Statistical learning and language acquisition* (pp. 337–364) Berlin: Mouton de Gruyter.

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., & Berry, D. C. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin and Review, 4*, 3–23.

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychological Research, 69*, 338–351.

Dulany, D. E., Carlson, R., & Dewey, G. (1984). A case of syntactical learning and judgement: How concrete and how abstract? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 113*, 541–555.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27,* 141–172.

Erdelyi, M., & Becker, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology,* 6, 159–171.

Faretta-Stutenberg, M., & Morgan-Short, K. (2011). Learning without awareness reconsidered: A replication of Williams (2005). In G. Granena, J. Koeth, S. Lee-Ellis, A. Lukyanchenko, G. Prieto Botana, & E. Rhoades (Eds.), *Selected proceedings of the 2010 Second Language Research Forum* (pp. 18–28). Somerville, MA: Cascadilla Proceedings Project.

Hama, M., & Leow, R. P. (2010). Learning without awareness revisited: Extending Williams (2005). *Studies in Second Language Acquisition, 32*, 465–491.

Hamrick, P. (2012, October). *Associative learning supports early phases of adult L2 syntactic development: Behavioral and computational evidence*. Poster presented at the Second Language Research Forum, Pittsburgh, PA.

Hamrick, P. (2013). *A role for chunk formation in statistical learning of second language syntax*. Manuscript submitted for publication.

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

Hamrick, P., & Rebuschat, P. (in press). Frequency effects, learning conditions, and the development of implicit and explicit lexical knowledge. In J. Connor-Linton & L. Amoroso (Eds.), *Measured language: Quantitative approaches to acquisition, assessment, processing and variation*. Washington, DC: Georgetown University Press.

Hamrick, P., & Sachs, R. (2013). *The need for appropriately trained control groups in applied linguistics research*. Manuscript in preparation.

Leow, R. P. (1997). Attention, awareness, and foreign language behavior. *Language Learning, 47*, 467–505.

Leow, R. P. (1998). Toward operationalizing the process of attention in SLA: Evidence for Tomlin and Villa's (1994) fine-grained analysis of attention. *Applied Psycholinguistics, 19*, 133–159.

Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: 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. H. C., & Williams, J. N. (2011a). Constraints on implicit learning of grammatical form-meaning connections. *Language Learning, 62*, 634–662.

Leung, J. H. C., & Williams, J. N. (2011b). The implicit learning of mappings between forms and contextually-derived meanings. *Studies in Second Language Acquisition, 33*, 33–55.

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

Reber, R., & Perruchet, P. (2003). The use of control groups in artificial grammar learning. *The Quarterly Journal of Experimental Psychology*, 56A, 97–115.

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

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

Rebuschat, P., Hamrick, P., Sachs, R., Riestenberg, K., & Ziegler, N. (2013). *Awareness, reactivity, and subjective measures*. Manuscript in preparation.

Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics, 33*, 829–856.

Rebuschat, P., & Williams, J. N. (2013). *Explicit learning and the development of implicit L2 knowledge*. Manuscript in preparation.

Reingold, E. M., & Merikle, P. M. (1990). On the inter-relatedness of theory and measurement in the study of unconscious processes. *Mind & Language, 5*, 10–28.

Robinson, P. (1995). Attention, memory and the 'noticing' hypothesis. *Language Learning, 45*, 283–331.

Robinson, P. (2003). Attention and memory during SLA. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 630–678). Oxford: Blackwell.

Rosa, E., & Leow, R. P. (2004). Awareness, different learning conditions, and L2 development. *Applied Psycholinguistics, 25*, 269–292.

Rosa, E., & O'Neill, M. D. (1999). Explicitness, intake, and the issue of awareness. *Studies in Second Language Acquisition*, *21*, 511–556.

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

Sachs, R., & Suh, B.-R. (2007). Textually enhanced recasts, learner awareness, and L2 outcomes in synchronous computer-mediated interaction. In A. Mackey (Ed.), *Conversational interaction and second language acquisition: A series of empirical studies* (pp. 197–227). Oxford: Oxford University Press.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11,* 129–158.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics, 13,* 206–226.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review—Revue de l'AILA, 11*, 11–26.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu, HI: University of Hawai'i, National Foreign Language Resource Center.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.

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

Tagarelli, K., Borges Mota, M., & Rebuschat, P. (2011). The role of working memory in 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.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, *16*, 183–203.

Vokey, J. R., & Higham, P. A. (2005). Abstract analogies and positive transfer in artificial grammar learning. *Canadian Journal of Experimental Psychology, 59*, 54–61.

Williams, J. N. (2004). Implicit learning of form-meaning connections. In B. VanPatten, J. Williams, S. Rott, & M. Overstreet (Eds.), *Form-meaning connections in second language acquisition* (pp. 203–218). Mahwah, NJ: Lawrence Erlbaum.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition, 27*, 269–304.

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, UK: Emerald Press.