Brain and Mind Institute Researchers'
Publications

Brain and Mind Institute

Implicit and explicit contributions to statistical learning.

Laura J Batterink
Northwestern University

Paul J Reber
Northwestern University

Helen J Neville
University of Oregon

Ken A Paller
Northwestern University

Implicit and explicit contributions to statistical learning

Laura J. Batterink[1], **Paul J. Reber**[1], **Helen J. Neville**[2], and **Ken A. Paller**[1]

[1]Northwestern University

[2]University of Oregon

Abstract

Statistical learning allows learners to detect regularities in the environment and appears to emerge automatically as a consequence of experience. Statistical learning paradigms bear many similarities to those of artificial grammar learning and other types of implicit learning. However, whether learning effects in statistical learning tasks are driven by implicit knowledge has not been thoroughly examined. The present study addressed this gap by examining the role of implicit and explicit knowledge within the context of a typical auditory statistical learning paradigm. Learners were exposed to a continuous stream of repeating nonsense words. Learning was tested (a) directly via a forced-choice recognition test combined with a remember/know procedure and (b) indirectly through a novel reaction time (RT) test. Behavior and brain potentials revealed statistical learning effects with both tests. On the recognition test, accurate responses were associated with subjective feelings of stronger recollection, and learned nonsense words relative to nonword foils elicited an enhanced late positive potential indicative of explicit knowledge. On the RT test, both RTs and P300 amplitudes differed as a function of syllable position, reflecting facilitation attributable to statistical learning. Explicit stimulus recognition did not correlate with RT or P300 effects on the RT test. These results provide evidence that explicit knowledge is accrued during statistical learning, while bringing out the possibility that dissociable implicit representations are acquired in parallel. The commonly used recognition measure primarily reflects explicit knowledge, and thus may underestimate the total amount of knowledge produced by statistical learning. Indirect measures may be more sensitive indices of learning, capturing knowledge above and beyond what is reflected by recognition accuracy.

Keywords

statistical learning; implicit learning; implicit memory; explicit memory; event-related potentials

Corresponding Author: Laura Batterink, Department of Psychology, 2029 Sheridan Road, Northwestern University, Evanston IL 60208, United States, lbatterink@northwestern.edu.

Introduction

Statistical learning refers to the process of extracting subtle patterns in the environment. This type of learning was first reported in 8-month-old infants, who were briefly exposed to a continuous stream of repeating three-syllable nonsense words. Following exposure, infants showed sensitivity to the difference between the three-syllable sequences and foil sequences made up of the same syllables recombined in a different order, demonstrating that they were able to use the statistics of the input stream to discover word boundaries in connected speech (Saffran, Aslin & Newport, 1996). This finding revolutionized thinking on language acquisition by showing that humans can use generalized statistical procedures to acquire language (Bates & Elman, 1996; Seidenberg, 1997).

Since this seminal study, subsequent research has shown that statistical learning can also be observed in older children and adults (e.g., Saffran et al., 1997, 1999, 2002; Fiser & Aslin, 2001, 2002; Turk-Browne et al., 2005). In a typical auditory statistical learning experiment run in adults, learners are exposed to a stream of repeating three-syllable nonsense words, as in Saffran and colleagues' original infant study. Learning is then assessed using a forced-choice recognition test. On each trial, learners are presented with a pair of stimuli: a nonsense word from the exposure stream is played together with a nonword foil composed of syllables from the speech stream combined in a novel order. Learners are asked to judge which stimulus sounds more familiar based upon the initial familiarization stream. Statistical learning is inferred if performance on this recognition measure is greater than chance.

An important feature of statistical learning is that it can occur in the absence of instruction or conscious attempts to extract the pattern, such as when stimuli are presented passively without any explicit task (e.g., Saffran et al., 1999; Fiser & Aslin, 2001, 2002; Toro, Sinnett & Soto-Faraco, 2005) or when participants are engaged in an unrelated cover task (Saffran et al., 1997, Turk-Browne et al., 2005, 2009). In addition, participants in statistical learning studies seem to have little explicit knowledge of the underlying statistical structure of the stimuli when assessed during debriefing (e.g., Conway & Christiansen, 2005; Turk-Browne et al., 2005). These results led researchers to describe statistical learning as occurring "incidentally" (Saffran et al., 1997), "involuntarily" (Fiser & Aslin, 2001), "automatically" (Fiser & Aslin, 2002), "without intent or awareness" (Turk-Browne et al., 2005), and "as a byproduct of mere exposure" (Saffran et al., 1999).

Statistical learning bears some similarity to *implicit learning*, a term coined by Art S. Reber (1967) and defined as "the capacity to learn without awareness of the products of learning" (Frensch & Runger, 2003). Paradigms used to study implicit learning include the artificial grammar learning (AGL) task (A.S. Reber, 1967) and the serial reaction time (SRT) task (Nissen & Bullemer, 1987). Learning in these tasks is typically measured indirectly, without making direct reference to prior studied items. In the AGL task, participants memorize letter strings generated by a grammatical rule system, and are then asked to decide whether new strings either conform to or violate the grammar. Above-chance classification performance is taken as evidence that participants have successfully acquired the underlying grammar. In the SRT task, participants respond to visual cues that contain a hidden repeating sequence. Participants eventually respond more quickly and accurately to sequential trials than to

random trials, indicating that they have learned the sequence. Thus, as in statistical learning, participants in implicit learning experiments are passively exposed to material that contains a hidden, repetitive structure. Learning proceeds as a consequence of exposure to positive examples, and in the absence of feedback or explicit instruction. In addition, both statistical learning and implicit learning are thought to be domain-general phenomena (e.g., Kirkham et al., 2002; Thiessen, 2011; Conway & Christiansen, 2005; Manza & A. S. Reber, 1997). The similarities between statistical learning and implicit learning have led some investigators to propose (or tacitly assume) that statistical learning and implicit learning arise due to the same general mechanism (e.g., Perruchet & Pacton, 2006; Conway & Christiansen, 2005; Turk-Browne et al., 2005).

In contrast to the statistical learning literature, the literature on implicit learning has focused on the nature of the representations formed during learning. These studies have sought to address whether the knowledge produced by implicit learning paradigms such as the AGL and SRT tasks is conscious (explicit) or unconscious (implicit). The use of confidence scales has been helpful in this regard. According to one widely accepted framework (Dienes & Berry, 1997), knowledge is implicit when participants lack meta-knowledge of what they have learned, either because they believe they are guessing when in fact they are above chance on a direct test of memory (*the guessing criterion*), or because their confidence is unrelated to their accuracy (*the zero-correlation criterion*). Thus, if participants perform above chance on a task when they claim to be guessing, or if they are no more confident when making correct responses compared to incorrect ones, knowledge is inferred to be implicit. In contrast, if participants perform above chance on the task, but their accuracy on guess responses is not higher than chance and/or they express greater confidence for correct responses compared to incorrect ones, knowledge is inferred to be explicit. These criteria apply only to *judgment* knowledge, defined as the ability to recognize whether a particular test item has the same structure as training items (Dienes & Scott, 2005). Judgment knowledge is distinct from *structural* knowledge, which is knowledge of the underlying structure of training materials and/or knowledge of the training items themselves. Judgment knowledge can be conscious even if structural knowledge is unconscious. In the present paper, we use the term "implicit knowledge" to refer to implicit judgment knowledge, as determined by the criteria of Dienes and Berry (1997).

Whether learning in AGL and SRT paradigms depends upon implicit knowledge has been a source of major contention in the literature. Original accounts of AGL concluded that learning in this paradigm is driven by the unconscious abstraction of information from the environment (e.g., A.S. Reber, 1967, 1976). According to this proposal, knowledge produced during the training phase was not accessible to awareness—participants acquired knowledge without realizing that they had acquired it. A number of subsequent studies supported this conclusion by showing that confidence ratings did not differ between correct and incorrect trials and that classification accuracy was better than chance even when participants claim to be guessing, collectively providing evidence of implicit judgment knowledge (Dienes et al., 1995; Dienes & Altmann, 1997; Tunney & Altmann, 2001; Scott & Dienes, 2008). Similarly, in the SRT task, participants often show robust learning as measured by performance while simultaneously exhibiting poor explicit recall or recognition

of the sequence, leading to the conclusion that sequence knowledge is implicit (e.g., P. J. Reber & Squire, 1994; Curran, 1997a, 1997b; Willingham & Goedert-Eschamann, 1999). Studies in amnesic patients provide additional support for this idea. Amnesic patients have been found to show intact performance on both the AGL as well as the SRT task, despite exhibiting greatly impoverished explicit memory (Knowlton, Ramus & Squire, 1992; Knowlton & Squire, 1994, 1996; P. J. Reber & Squire, 1994). These results indicate that explicit knowledge of the training materials or underlying sequence is not needed to support performance on these tasks. It is important to note that implicit knowledge does not necessarily consist exclusively of abstract rule knowledge, as originally proposed by A. S. Reber (1967, 1976). Concrete, item-specific knowledge, such as memory of specific letter strings, can be acquired independently of declarative memory, and this implicit knowledge can also support classification performance on the AGL task (Knowlton & Squire, 1996).

However, a number of arguments have been made against the "two-systems" view that performance on these implicit learning paradigms reflects implicit knowledge dissociable from explicit knowledge. One common argument is that implicit learning paradigms frequently produce explicit knowledge in healthy adults, and this explicit knowledge can also account for performance on these tasks. For instance, on the AGL task, healthy learners may form explicit memories for some of the instances or chunks, which can then be used to guide classification decisions (Perruchet, Gallego & Savy, 1990; Perruchet & Pacteau, 1990; Servan-Schreiber & Anderson, 1990). In principle, structural knowledge of certain rules (e.g., the knowledge that an "M" can start a string) may also be conscious, and this knowledge may also be recruited during the classification test (Dienes & Scott, 2005). A second commonly-raised objection is that studies demonstrating implicit learning often fails to adequately assess awareness of knowledge (e.g., Shanks & St. John, 1994). For example, sequence knowledge in SRT tasks may be accessible through certain free-generation and recognition tasks, raising questions about whether this knowledge can really be characterized as unconscious (Shanks & Johnstone, 1999). Finally, a number of authors have challenged the two-systems view from a logic standpoint, arguing that the dissociation between performance and awareness can be accounted for without invoking separate implicit and explicit learning systems (Destrebecqz & Cleeremans, 2003; Shanks, Wilkinson, & Channon, 2003; Shanks & Perruchet, 2002; Kinder & Shanks, 2003). For example, representations formed during learning may be of insufficient quality to support conscious awareness, but still be adequate to influence behavior (Destrebeczq & Cleeremans, 2003). Consistent with this idea, a recent modeling approach has shown that imperfect memories for training exemplars would be sufficient to support classification judgments performance on the AGL task, despite being too noisy to support retrieval for recall (Jamieson & Mewhort, 2009a). The same model was also applied successfully to account for performance on the SRT, demonstrating that very local memory for events can be used to speed responding while being insufficient to support retrieval of the sequence (Jamieson & Mewhort, 2009b). According to this approach, performance on the AGL and SRT paradigms can be explained by the same principles used for explicit-memory tasks.

As demonstrated by these different lines of evidence, it is challenging to unequivocally demonstrate the existence of implicit learning in healthy adults, in whom both implicit and explicit learning systems are fully functioning. Nonetheless, neuropsychological evidence

from amnesic patients (Knowlton, Ramus & Squire, 1992; Knowlton & Squire, 1994, 1996; P. J. Reber & Squire, 1994) as well as neural dissociations between implicit and explicit learning systems in healthy subjects (e.g., Poldrack et al., 2001; Foerde et al., 2006; Liebermann et al., 2004) offer strong general support for the two-systems view. On the AGL task, it has been shown that explicit and implicit knowledge can be dissociated, with both types of knowledge capable of supporting performance (Higham, 1997; Vokey & Brooks, 1992; Meulemans & Van der Linden, 1997; Liebermann et al., 2004). Similarly, by manipulating participants' explicit awareness of the sequence, SRT studies have demonstrated that unconscious procedural learning occurs whether or not it is accompanied by explicit sequence knowledge; the development of explicit knowledge simply occurs in parallel (Willingham & Goedert-Eschmann, 1999; Willingham, Salidis, & Gabrieli, 2002). This has also been found to apply to contextual cueing, in which explicit memorization of visual scene information engages neural processes beyond those required for the implicit learning of target locations (Westerberg et al., 2011). It has also been shown that explicit knowledge does not directly contribute to task performance in normal participants in a variant of the SRT, the Serial Interception Sequence Learning (SISL) task (Sanchez & P. J. Reber, 2013).

Taken together, the results indicate that implicit learning paradigms such as the AGL and SRT frequently result in the parallel acquisition of both implicit and explicit knowledge in healthy adult learners. Nonetheless, the development of explicit knowledge is optional and can often be dissociated from implicit knowledge. By extension, we hypothesize that statistical learning, to the extent that it resembles implicit learning, may produce implicit knowledge optionally accompanied by explicit knowledge.

This hypothesis has not been thoroughly examined, as the focus of most statistical learning studies has been on the properties of the learning process, rather than on the products of learning. However, within the last several years several researchers have begun to investigate the nature of the representations formed during statistical learning. Kim and colleagues (2009) exposed participants to a stream of structured visual stimuli, and then used an indirect reaction time (RT) task to assess learning and a direct item-matching test to assess participants' awareness of learning. Participants in this study showed RT effects in the absence of explicit knowledge, as indicated by chance performance on the matching test, leading the authors to conclude that statistical learning involves implicit learning mechanisms. In contrast to these conclusions, Bertels and colleagues (2012, 2013) showed that participants successfully scored above chance on an easier, putatively more sensitive version of the explicit matching test, suggesting that participants' sequence knowledge is at least partially available to consciousness. Nonetheless, participants also performed above chance on this task even when they claimed to be guessing, suggesting that performance was at least partly based on implicit knowledge. Finally, Franco and colleagues (2011) used Jacoby's (1991) Process Dissociation Procedure to examine whether participants' recognition of training items is driven entirely by familiarity, or whether conscious recollection also contributes. Learners were exposed to two different artificial speech streams, and then completed an "inclusion" test, in which they were asked to distinguish between words from either stream and new words, and an "exclusion" test, in which they were instructed to respond to only the words from the first (or second) stream. Successful

performance on the exclusion test is assumed to be based on conscious recollection, as a mere feeling of familiarity would lead participants to respond to items belonging to both streams. Franco and colleagues (2011) found that learners could successfully differentiate items from the two different streams, providing evidence that statistical learning produces representations that can be consciously controlled. Although these results are somewhat mixed, taken together they suggest that statistical learning produces representations that are at least partially explicit in nature.

Building upon this fledgling literature, we adopted principles established in the implicit learning literature to examine the role of explicit knowledge within the context of a typical auditory statistical learning paradigm (Saffran et al., 1996, 1996b, 1997). One possibility is that performance on tasks used to assess auditory statistical learning can be entirely accounted for by explicit knowledge, resembling some accounts of implicit learning that have been proposed to account for AGL and SRT learning effects (Destrebecqz & Cleeremans, 2003; Shanks, Wilkinson, & Channon, 2003; Shanks & Perruchet, 2002; Kinder & Shanks, 2003). Alternatively, both implicit and explicit knowledge may contribute to statistical learning, such that explicit knowledge alone is insufficient to account for observed learning effects. To distinguish between these two possibilities, we used both direct and indirect tests of memory to characterize the knowledge produced during statistical learning. Although both tasks can potentially be sensitive to both implicit and explicit influences, direct tests of memory make reference to previously studied items, whereas indirect tests surreptitiously measure knowledge without requiring participants to make a decision about whether they have previously encountered an item. Thus, direct tests of memory are generally more sensitive to explicit knowledge, whereas indirect tests of memory are generally more sensitive to implicit knowledge. To the extent that statistical learning generates implicit knowledge, using only a direct measure of learning (as is typical of most statistical learning studies) may run the risk of underestimating the total amount of knowledge that has been acquired. The combination of direct and indirect measures can therefore provide a more comprehensive picture of the knowledge acquired during statistical learning.

To outline the present study, we employed a typical auditory statistical learning paradigm in which participants are exposed to a continuous stream of nonsense words (e.g., Saffran et al., 1996, Saffran et al., 1997). We then assessed learning in two ways: (a) directly, through a forced-choice recognition task, the most common way of assessing auditory statistical learning and (b) indirectly, through a reaction-time-based target-detection task. This target detection task, adapted from a paradigm used previously in the visual statistical learning literature (Kim et al., 2009; Olson and Chun, 2001; Turk-Browne et al., 2005), has not previously been applied to auditory statistical learning. In Experiment 1, we acquired behavioral data during these two tasks. In Experiment 2, we employed an additional behavioral procedure, the remember/know task, during recognition testing, which allowed us to apply the criteria of Dienes and Berry (1997) to assess learners' awareness of their knowledge. We also recorded event-related potentials (ERPs) to shed additional insight into the nature of the representations produced by statistical learning. Combined results from these two experiments supported the hypothesis that statistical learning produces both implicit and explicit knowledge.

Experiment 1

Experiment 1 was a behavioral study designed to confirm that auditory statistical learning can be observed through both a direct and an indirect test of memory. We expected that participants would score above chance on the recognition task (the direct test) and would show priming on the target detection task (the indirect test), as reflected by RTs. We also assessed participants' overall subjective confidence on the recognition task as an initial test of whether this measure reflects implicit or explicit processes. According to the zero-correlation criterion—which has usually been applied to trial-by-trial confidence ratings—confidence and accuracy should be uncorrelated if performance is driven by implicit memory. If confidence and accuracy are positively correlated, this would violate the principle behind this criterion, providing evidence that recognition judgments are at least partially supported by explicit memory. Finally, as a secondary question, we examined correlations between performance on the recognition and target detection tasks. Evidence that performance on these two tasks is uncorrelated would be consistent with the idea that statistical learning produces both implicit and explicit knowledge.

Materials and Methods

Participants—Twenty-four native English speakers (12 women) were recruited at the University of Oregon to participate in the experiment. Participants were between 18 and 31 years old (M = 24.0 years, SD = 4.6 years) and had no history of neurological problems. To examine whether statistical learning differs under intentional versus incidental learning conditions, participants were randomly assigned to an implicit (*n* = 12) or explicit (*n* = 12) instruction condition (described in greater detail below). However, instruction condition did not have a significant effect on any dependent measure, and thus the main results described below were collapsed across participants from both groups. Participants were paid $10/hr.

Stimuli—For the learning phase, stimuli and experimental parameters were modeled after those used in previous auditory statistical learning studies (e.g., Saffran, Newport & Aslin, 1996; Saffran et al., 1997). This stimulus set consisted of 11 syllables combined to create six trisyllabic nonsense words, henceforth called words (babupu, bupada, dutaba, patubi, pidabu, tutibu). Some members of the syllable inventory occurred in more words than others in order to ensure varying transitional probabilities within the words themselves, as in natural language. A speech synthesizer was used to generate a continuous speech stream composed of the six words at a rate of approximately 208 syllables per minute, approximating the rates used in previous auditory statistical learning experiments conducted in adults (Saffran et al., 1996b; Saffran et al., 1997). Each word was repeated 300 times in pseudorandom order, with the restriction that the same word never occurred consecutively. Because the speech stream contained no pauses or other acoustic indications of word onsets, the only cues to word boundaries were statistical in nature (either transitional probabilities, which were higher within words than across word boundaries, or frequency of co-occurrence, which were higher for the three syllable sequences within words than across words; cf. Saffran et al., 1996b; Saffran et al., 1997; Aslin, Saffran & Newport, 1998). The speech stream was divided into 3 equal blocks, each one approximately 8 minutes in length.

Stimuli and experimental parameters for the recognition test also followed previous auditory statistical learning experiments conducted in adults (Saffran et al., 1996b; Saffran et al., 1997). Six nonword foils were created (batabu, bipabu, butipa, dupitu, pubada, tubuda). The nonwords consisted of syllables from the language's syllable inventory that never followed each other in the speech stream, even across word boundaries. Participants were tested on nonword foils rather than part-word foils, which consist of two syllables from a word plus an additional syllable, as discrimination accuracy is typically higher when nonwords are used (Safffran et al., 1996b). By using nonwords, we hoped to obtain higher levels of recognition accuracy and increase the sensitivity of this explicit measure, which in turn should yield greater power to detect potential correlations between recognition and priming. In a departure from early auditory statistical learning studies (e.g., Saffran, Newport & Aslin, 1996b; Saffran et al., 1997; Sanders et al., 2002), the frequency of individual syllables across words and nonword foils were also matched. This represents an improvement on the original stimulus streams, which failed to match the number of occurrences of each individual syllable across words and nonword foils. In past studies using these stimuli, learners' recognition of a single highly familiar syllable that is represented more frequently in words than nonword foils—rather than sensitivity to the distribution of syllables across time—could in principle produce above-chance performance on the recognition task. In addition, we confirmed that any preference for words over nonwords could not be attributed to systematic differences between items by running a group of control participants ($n = 11$), who completed the recognition task without prior exposure to the speech stream. Control participants' preference for the words over the nonwords was not reliably above chance (50.5%, $t(10) = 0.17$, $p = 0.87$), indicating that above-chance performance on the recognition task cannot be attributed to item differences between words and nonwords.

Finally, for the speeded target detection task, 33 separate speech streams were created with the same speech synthesizer used to create stimuli in the learning phase. Each stream consisted of two repetitions of each of the six nonsense words, concatenated together in pseudorandom order. The speech streams for the target detection task were produced at a somewhat slower rate than the original speech streams (approximately 144 syllables per minute). This moderate rate was chosen to ensure that the task would be both feasible and challenging, in order to provide a direct measure of online speeded processing. In order to compute RTs to target syllables, target syllables onsets were coded by three trained raters using both auditory cues and visual inspection of sound spectrographs. Any discrepancy greater than 20 ms among one or more raters was resolved by a fourth independent rater.

Procedure—Participants in both instruction conditions were exposed to the same auditory streams. Participants in the implicit condition were instructed to listen to the auditory stimuli. Participants in the explicit condition were informed that that they would be listening to a "nonsense" language that contained words, but no meanings or grammar. They were informed that their task was to figure out where each word began and ended, and that they would be tested on their knowledge of the words at the end of the experiment. They were not given any information about the length or structure of the words or how many words the language contained. All auditory stimuli were played at a comfortable listening level from speakers mounted on either side of the participant.

After finishing the listening phase of the experiment, participants in the implicit condition were informed that the auditory stimuli that they had just listened to were actually words in a nonsense language. All participants then completed a forced-choice recognition task. Each trial included a word and a nonword foil. The task was to indicate which of the two sound strings sounded more like a word from the language. Each of the six words and six nonword foils were paired exhaustively for a total of 36 trials. In half of the trials the word was presented first while in the other half the nonword foil was presented first; presentation order for each individual trial was counterbalanced across subjects. Each trial began with the presentation of a fixation cross. After 1000 ms, the first word was presented. The second word was presented 1500 ms after the onset of the first word. Individual word duration ranged from 800–900 ms. Participants' overall subjective confidence in performing this task was assessed in a post-experiment questionnaire, in which they were asked to rate on a 1–10 scale how often they felt confident that their response was correct.

Finally, participants completed the speeded target detection task, in which they detected target syllables within a continuous speech stream made up of the six words. Both RT and accuracy were emphasized. Each of the 11 syllables of the language's syllable inventory (ba, bi, bu, da, du, pa, pi, pu, ta, ti and tu) served as the target syllable three times, for a total of 33 streams. Prior to the presentation of each of the 33 streams, participants were instructed to detect a specific target syllable (e.g., "ba") within the continuous speech stream. Each stream contained between 2 to 8 target syllables, depending upon which syllable served as the target. Across all 33 streams there was a total of 36 "trials" in each of the three syllable conditions (word-initial, word-middle, and word-final). A "trial" in this sense refers to the presentation of a target syllable within the continuous speech stream. It was expected that RTs would be fastest to syllable targets in the final position of a word, with word-initial and word-middle targets eliciting slowest and intermediate RTs, respectively. Faster RTs to later syllables would reflect priming effects elicited by the presentation of earlier syllables within a word, and would be consistent with behavioral RT effects reported in the visual statistical learning literature using a similar paradigm (Kim et al., 2009; Olson and Chun, 2001; Turk-Browne et al., 2005). Before each of the syllable streams was presented, participants pressed "Enter" to listen to a sample of the target syllable. The syllable stream was then initiated. The order of the 33 streams was randomized for each participant. The duration of each stimulus stream was approximately 15 s, with an average SOA between syllables of approximately 400 ms. The interval between individual syllables was jittered due to natural variability in the speech streams created by the speech synthesizer.

Analysis—For the target detection task, median RTs were calculated for each syllable condition (word-initial, word-middle, and word-final). Only responses that occurred between 150 and 1200 ms after target onset were included; all other responses were considered to be false alarms. RTs were analyzed using a repeated-measures ANOVA with syllable position (initial, middle, final) as a within-subjects factor and instruction condition (implicit, explicit) as a between-subjects factor. Planned contrasts were used to examine whether RTs decreased linearly as a function of syllable position.

Results

Recognition Task—Mean accuracy across all participants was moderate (61.3%, *SD* = 8.0%) but significantly better than the chance level of 50% [$t(23) = 6.96$, $p < 0.0001$; *Cohen's d* = 2.90]. There was no significant difference in performance between implicit and explicit instruction groups [implicit group: 59.0%, SD = 5.7%; explicit group: 63.7%, SD = 9.4%, $t(22) = 1.46$, $p = 0.16$; *Cohen's d* = 0.60]. However, participants in the explicit instruction group had significantly higher subjective confidence ratings than participants in the implicit group [implicit group: 4.8/10; explicit group: 6.75/10; $t(22) = 2.56$, p = 0.017; *Cohen's d* = 1.10). Recognition accuracy significantly correlated with subjective confidence across participants, as those with higher recognition accuracy were more confident ($r = 0.48$, $p = 0.018$).

Target Detection Task—RTs are plotted in Figure 1A. Across all participants, RTs showed the predicted decrease for later syllable positions [Position effect: $F(2,44) = 28.6$, $p < 0.001$; $\eta^2_p = 0.57$; linear contrast: $F(1,22) = 38.2$, $p < 0.0001$; $\eta^2_p = 0.64$]. Planned contrasts revealed that RTs were significantly faster for the final position relative to the middle position [$F(1,22) = 50.8$, $p < 0.0001$; $\eta^2_p = 0.70$], but not significantly different between the middle and initial position [$F(1,22) = 0.90$, $p = 0.35$; $\eta^2_p = 0.039$]. There was no significant difference in RT effects between implicit and explicit participants [Group x Syllable Position: $F(2,44) = 0.15$; $p = 0.845$; $\eta^2_p = 0.007$].

Because the target detection task necessarily involves additional exposure to the stimulus stream, one potential concern is that performance may be driven by learning that occurred during the target detection task itself, rather than during the initial exposure period. If this were the case, one would expect the magnitude of the RT effect to be larger in the second half of the task. Contrary to this idea, there was no significant difference between the RT effect in the first and second half (Task Half x Syllable Position: $F(2,46) = 1.75$, $p = 0.19$; $\eta^2_p = 0.071$). Even if some additional learning did occur during the task, this learning would also be statistical in nature, as the statistical probabilities between syllables were the only cue predicting the upcoming stimuli.

Following previous statistical learning studies (e.g., Saffran, Newport & Aslin, 1996; Saffran et al., 1997; Sanders et al., 2002), we used stimuli in which some syllables are represented more frequently across words than others, in order to ensure varying transitional probabilities between syllables within words. Therefore, there is a potential confound between frequency of a given syllable and syllable position. In particular, the syllable "bu" was represented more frequently than the other ten syllables, and occurred more often in the final position compared to the first two. To examine whether the higher frequency of this syllable may have driven observed RT priming effects (i.e., faster RTs for targets occurring in later syllable positions), we removed all instances of "bu" from analysis. RT priming effects remained robust (Position effect: $F(2,44) = 13.8$, $p < 0.001$; $\eta^2_p = 0.39$; linear contrast: $F(1,22) = 19.0$, $p < 0.001$; $\eta^2_p = 0.46$), demonstrating that unbalanced syllable representation across words cannot account for our observed effects.

Relationship between Recognition and RT—Correlations were calculated between each participant's recognition score and their RT priming effect, computed as the RT difference between initial position and final position syllables ($RT_1 – RT_3$). As shown in Figure 1B, there was no significant correlation between these two measures ($r = 0.07$, $p = 0.75$, 95% CI for $r = –0.34$ to 0.46).

Discussion

We found that statistical learning not only produced above-chance discrimination on the recognition measure, but also resulted in robust RT effects as assessed during the speeded target detection task. Auditory statistical learning can thus be observed using both direct and indirect measures. Participants' subjective confidence significantly correlated with their performance on the recognition task, providing evidence that explicit mechanisms contribute to accurate recognition judgments. As a secondary point of discussion, the increase in detection speed on the target detection task did not correlate with explicit recognition. Although we cannot rule out that this lack of correlation was due to variability in the measures or to low statistical power, this finding leaves open the possibility that implicit knowledge of the statistical structure was accrued in parallel with explicit knowledge during exposure to the speech streams.

These results are consistent with the notion that the recognition measure commonly used in statistical learning paradigms at least partially reflects explicit memory. In Experiment 2, we built upon these results by incorporating a trial-by-trial measure of memory experiences in the recognition task, so that we could test more precisely whether performance is driven by explicit memory. Data collected in Experiment 2 also served to evaluate the reliability of the apparent dissociation between the recognition and target detection tasks. In addition to behavioral measures, ERPs were recorded to examine the neural mechanisms recruited to support performance, so as to test for functional dissociations during the two tasks. Based on results from Experiment 1, we hypothesized that explicit knowledge from statistical learning would again be evident during the recognition task. As a secondary point, we also predicted that the recognition and target detection tasks would again show dissociations across participants, as would be predicted if statistical learning produces both implicit and explicit knowledge.

Experiment 2

In principle, a forced-choice recognition measure such as the one used in conventional tests of statistical learning may rely upon either implicit or explicit knowledge (Paller, Voss & Boehm 2007; Voss, Lucas & Paller, 2009; Voss et al., 2008; Voss & Paller, 2009). Thus, to examine whether recognition judgments are primarily supported by implicit or explicit memory, we adopted a remember/know procedure for the recognition task, in which participants were asked to report on experiential aspects of memory retrieval on each trial. In this procedure, *remember* indicates confidence based on retrieving specific information from the learning episode, *familiar* indicates a vague feeling of familiarity with no specific retrieval, and *guess* indicates no confidence in the selection. Note that familiar responses reflect explicit judgment knowledge, in that the participant shows some degree of confidence in the correctness of his or her response, but could potentially reflect implicit

structural knowledge, in that the knowledge used to make the response may be non-verbalizable. If recognition judgments are largely supported by explicit memory retrieval, we should expect to see highest accuracy for "remember" responses, moderate accuracy for "familiar" responses, and lowest accuracy for "guess" responses. In contrast, if performance on this task is supported by implicit (judgment) knowledge, awareness of memory retrieval should not differ between correct and incorrect responses (the zero-correlation criterion) and/or should be above chance when participants claim to be guessing (the guessing criterion).

We also recorded brain potentials during both the recognition and target detection tasks in order to examine the nature of the processes used to support performance in these two tasks. For the recognition task, we focused on the LPC, a positive-going ERP modulation with an onset of approximately 400–500 ms post-stimulus-onset that has been specifically linked to recollection (Paller & Kutas, 1992; Rugg & Curran, 2007). The LPC is typically observed in contrasts between old and new items in recognition tests (e.g., Rugg et al., 1998; Rugg & Curran, 2007; Voss & Paller, 2008) and may reflect the amount of information recollected in response to a test item (Vilberg et al., 2006), with larger LPC amplitude indicating better or more detailed recollection. Therefore, we hypothesized that if recognition judgments are supported by explicit memory, learned words should elicit a larger LPC than nonword foils.

With respect to the target detection task, we focused on P300, a positive-going ERP component with a typical latency of approximately 250–500 ms post-stimulus-onset that is elicited during stimulus discrimination (Polich, 2007). Early studies using the two-stimulus oddball task demonstrated that discriminating a target stimulus from a stream of standards elicits a robust P300, with P300 amplitude correlating inversely with target probability (Duncan-Johnson and Donchin, 1977, 1982; Johnson and Donchin, 1982; Squires et al., 1976). One widely accepted theory proposes that P300 reflects the allocation of attentional resources to the target, which are engaged in order to update the current neural representation of the stimulus environment (Polich, 2003, 2007). When task demands and overall levels of attention and arousal are held constant, targets that are less probable or predictable elicit larger P300 effects, in keeping with the idea that unpredictable targets engage greater attentional resources. We therefore hypothesized that a reduced P300 should be elicited to predictable syllable targets (i.e. those that occur in later syllable positions), reflecting a facilitation in processing due to statistical learning. This pattern of results would converge with RT effects observed in Experiment 1 and provide additional evidence that statistical learning produces representations that can be rapidly recruited to support performance on an indirect test of memory. This P300 effect would also rule out an alternative explanation for the RT effect observed in Experiment 1: namely, that faster RTs are driven by a *greater* allocation of attentional resources, rather than reflecting a true facilitation in processing. That is, participants may have responded more quickly to third syllable targets than first syllable targets because they directed a greater amount of effortful, controlled processes to these targets (perhaps because the contextual information supplied by preceding syllables provided them with more time to engage top-down processes). Such an account would be inconsistent with the idea that knowledge acquired during statistical

learning leads to greater processing efficiency, enabling the brain to allocate limited-capacity resources to other ongoing tasks.

Finally, as a secondary hypothesis, we predicted that there would be no correlation between recognition and both RT and ERP indices of priming on the target detection task, consistent with the idea that the target detection task reflects implicit representations dissociable from explicit memory. In contrast, we expected to find correlations between RT and P300 effects on the target detection task, indicating that both behavioral and ERP measures of priming are driven by a common underlying mechanism.

Method

Participants—Twenty-five native English speakers (13 women) were recruited at the University of Oregon to participate in the experiment. Participants were between 18 and 30 years old (M = 20.5 years, SD = 2.5 years), were right-handed, and had no history of neurological problems. Participants were randomly assigned to an implicit (n = 13) or explicit (n = 12) instruction condition. However, instruction condition did not have a significant impact on any dependent measure, and thus the main results described below were collapsed across participants from both groups. Participants earned course credits for their participation. Data from all 25 participants were included in all behavioral analyses and ERP analyses for the target detection task. However, three participants' EEG data were excluded from the recognition task because of excessive artifact, resulting in a final sample of 22 participants for ERP analyses related to the recognition task.

Stimuli—For the learning phase, speech streams were similar to those used in Experiment 1, with two minor exceptions. First, the speech stream was presented at a slightly faster rate (approximately 255 syllables/minute), in order to equate the total duration of exposure (21 minutes) with previous auditory statistical learning studies (Saffran et al., 1996; Saffran et al., 1997). Second, the speech stream included a total of 31 brief pitch changes. Each pitch change represented either a 20 Hz increase or decrease from the baseline frequency, and spanned four consecutive syllables. These pitch changes were introduced in order to provide participants with an unrelated cover task during the learning period and to ensure attention to the auditory stimuli. Pitch changes occurred randomly, rather than systematically on certain syllables, and thus could not provide a cue for segmentation.

For the Recognition task, the onsets of each word and nonword foil relative to the beginning of the sound file were identified both auditorily and through visual inspection of the audiogram and subsequently coded for ERP analysis. For the target detection task, the same onset times coded for RT analyses were used for the ERP analysis. Stimulus parameters (durations and inter-stimulus-intervals) were identical to those in Experiment 1.

Procedure—The procedure for Experiment 2 was similar to that in Experiment 1. At the beginning of the experiment, participants were fitted with an elastic EEG cap embedded with electrodes. All participants were then exposed to the auditory stimuli, with the same instructional manipulation as described for Experiment 1. Prior to the exposure phase, participants were informed that the speech stream contained occasional pitch changes and that they should press one button for low pitch changes and another for high pitch changes.

Test order was counterbalanced across participants, with approximately half of participants completing the recognition task and the others completing the target detection task first.

For the recognition task, participants gave two responses for each trial. First, they indicated which of the two sound strings was more familiar, as in Experiment 1. After selecting this response, they then reported on their awareness of memory retrieval. Specifically, they were instructed to respond "remember" if they felt confident in their choice and had a memory of the word based on the prior learning episode, "familiar" if they felt that one of the words was more familiar than the other, but did not have a specific memory for the word, and "guess" if they had no idea which stimulus was correct and felt as though they were being forced to choose one at random.

EEG Recording and Analysis—EEG was recorded at a sampling rate of 2048 Hz from 64 Ag/AgCl-tipped electrodes attached to an electrode cap using the 10/20 system. Recordings were made with the Active-Two system (Biosemi, Amsterdam, Netherlands), which does not require impedance measurements, an online reference, or gain adjustments. Additional electrodes were placed on the left and right mastoid, at the outer canthi of both eyes and below the right eye. Scalp signals were recorded relative to the Common Mode Sense (CMS) active electrode and then re-referenced off-line to the algebraic average of the left and right mastoid. Left and right horizontal eye channels were re-referenced to one another, and the vertical eye channel was re-referenced to FP1.

ERP analyses were carried out using EEGLAB (Delorme & Makeig, 2004). Data were down-sampled to 1024 Hz and then band-pass filtered from 0.1 to 40 Hz. Large or paroxysmal artifacts or movement artifacts were identified by visual inspection and removed from further analysis. Data were then submitted to an Independent Component Analysis (ICA), using the extended runica routine of EEGLAB software. Ocular and channel artifacts were identified from ICA scalp topographies and the component time series, and removed. ICA-cleaned data were then subjected to a manual artifact correction step to detect any residual or atypical ocular artifacts not removed completely with ICA. For a subset of subjects ($n = 3$), one or more channels were identified as bad, excluded from all ICA decompositions, and interpolated later. Finally, epochs time-locked to critical events were extracted and plotted from –100 to 1200 ms, and baseline corrected to a 100-ms prestimulus interval. In the recognition task, averages were time-locked to the onsets of word and nonword foils (approximate duration of spoken words = 800–1000 ms), whereas in the target detection task averages were time-locked to the onset of target syllables.

To investigate whether larger LPCs were elicited by words than nonword foils in the recognition task, mean LPC amplitudes to words and nonword foils were calculated for each participant. Two separate analyses were conducted, one that included all trials and a second that included only trials to which participants responded correctly (i.e., excluding incorrect responses). On the basis of previously published findings (Rugg et al., 1998; Rugg & Curran, 2007; Voss & Paller, 2008; Friedman & Johnson, 2000) and visual inspection of the waveforms, the LPC time interval was selected as 700–1000 ms poststimulus. Given that spoken words take some time to be presented, this interval is somewhat later than the typical interval selected by ERP recognition studies using visual stimuli. To increase the sensitivity

of this test, channels for LPC analysis were selected *a priori* according to where the LPC effect was expected to be maximal, and included only posterior channels (PO7, PO3, O1, Pz, POz, Oz, PO8, PO4, & O2). Mean amplitude values for these channels were averaged together and analyzed using a repeated-measures ANOVA with word class (word, nonword foil) as a within-subjects factor and instruction condition (implicit, explicit) as a between-subjects factor.

To investigate whether the amplitude of the P300 elicited by target syllables during the target detection task varied as a function of syllable position, mean P300 amplitudes to target syllables in the three syllable conditions (initial, middle, and final) were calculated for each participant. Only trials to which participants made a correct response within 1200 ms were included in the analysis. The P300 time-interval from 400 to 800 ms was selected on the basis of previous studies and on visual inspection of the data (Polich, 2007). Channels for P300 analyses were selected to include only central and posterior electrodes, as the P300 typically shows the largest distribution over parietal regions (Polich, 2007). Following our usual procedures (e.g., Batterink & Neville, 2013), amplitudes were averaged across neighboring electrodes to form nine electrode regions of interest (left anterior region: AF7, AF3, F7, F5, F3; left central region: FT7, FC5, FC3, T7, C5, C3; left posterior region: TP7, CP5, CP3, P7, P5, P3, PO7, PO3; midline anterior region: AFZ, F1, FZ, F2; midline central region: FC1, FCZ, FC2, C1, CZ, C2; midline posterior region: CP1, CPZ, CP2, P1, PZ, P2, POZ; right anterior region: AF4, AF8, F4, F6, F8; Right central region: FC4, FC6, FT8, C4, C6, T8; right posterior region: CP4, CP6, TP8, P4, P6, P8, PO4, PO8). Mean amplitude values of these nine electrode regions were initially submitted to a repeated-measures ANOVA, with syllable position (initial, middle, final), anterior-posterior axis (central, posterior) and left/right (left, midline, right) as within-subjects factors, and with instruction condition (implicit, explicit) as a between-subjects factor. Greenhouse-Geisser corrections were applied for factors with more than two levels.

For analyses with all trials in the recognition task, each participant contributed an average of 32 trials (range = 15–36) to each condition (word, nonword). For analyses for correct trials in the recognition task, each participant contributed an average of 20 trials (range = 11–29) to each condition. For analyses in the target detection task, each participant contributed an average of 29 trials (range = 16–35) to each of the three syllable conditions.

Results

Behavioral Results

__Learning Task:__ Overall, participants performed well on the pitch detection cover task. They detected 92% (SD = 8.0%) of the 31 pitch changes, with an average of 6.7 false alarms (SD = 8.3).

__Recognition Task:__ Behavioral results were generally similar to those from Experiment 1. Mean accuracy on the recognition task (mean = 58.7%, SD =11.8%) was slightly lower than in Experiment 1 though still significantly above chance [$t(24) = 3.67$, $p = 0.001$; *Cohen's d* = 1.50]. Again, there were no significant differences in performance between implicit and explicit groups [$t(23) = 0.70$, $p = 0.49$; *Cohen's d* = 0.29].

Across all participants, "remember" responses were the most accurate followed by "familiar" responses, with "guess" judgments showing the lowest degree of accuracy [Memory Judgment effect: $F(2,44) = 5.65$, $p = 0.009$; $\eta^2_p = 0.20$; linear contrast: $F(1,22) = 8.51$, $p = 0.008$; $\eta^2_p = 0.28$; Figure 2A). When participants claimed to be guessing, accuracy was not significantly above chance [mean = 49.4%, SD=23.5%; $t(22) = 0.13$, $p = 0.90$; *Cohen's d* =0.055]. Accuracy across these three metamemory responses was not significantly different between implicit and explicit participants [Group x Memory Judgment: $F(2, 42) = 0.046$, $p = 0.94$; $\eta^2_p = 0.002$], nor was the overall proportion of responses different between the two groups [$F(2,46) = 0.27$, $p = 0.76$; $\eta^2_p = 0.012$].

Target Detection Task: RTs showed a similar pattern as in Experiment 1 (Figure 2B). Across all participants, RTs were faster for syllables occurring in later positions [Position effect: $F(2,46) = 23.0$, $p < 0.0001$; $\eta^2_p = 0.50$; linear contrast: $F(1,23) = 22.6$, $p < 0.0001$; $\eta^2_p = 0.49$]. Planned contrasts revealed that there was a significant RT facilitation for the final position compared to the middle position [$F(1,23) = 35.3$, $p < 0.001$; $\eta^2_p = 0.61$], but RTs did not significantly differ between initial and middle positions [$F(1,23) = 1.70$, $p = 0.21$; $\eta^2_p = 0.069$]. Again, no significant difference in RT effects were found between implicit and explicit participants [Group x Syllable Position: $F(2,46) = 1.05$, $p = 0.35$; $\eta^2_p = 0.043$]. The RT effect did not significantly differ between the first and second half of the task [Task Half x Syllable position: $F(2,48) = 0.98$, $p = 0.38$; $\eta^2_p = 0.039$], and this effect was also nonsignificant when participants from both experiments were combined to increase power [Task Half x Syllable position: $F(2,96) = 1.50$, $p = 0.23$; $\eta^2_p = 0.030$]. As in Experiment 1, RT effects remained robust even after removal of the syllable "bu" from analysis (Position effect: $F(2,46) = 9.87$, $p = 0.001$; $\eta^2_p = 0.30$; linear contrast: $F(1,23) = 10.3$, $p = 0.004$; $\eta^2_p = 0.31$), demonstrating that the potential confound between frequency and syllable position cannot account for observed effects.

Relationship Between Recognition and RT Priming: As in Experiment 1, the across-subject correlation between recognition scores and the magnitude of the RT priming effect ($RT_1 - RT_3$) on the target detection task was nonsignificant ($r = 0.26$, $p = 0.20$, 95% CI for r = −0.15 to 0.60). However, this correlation may have been artificially inflated by three participants who showed no behavioral evidence of learning as assessed by either the recognition or the RT measure [criteria: 50% accuracy in recognition task, < 10 ms effect ($RT_1 - RT_3$) in target detection task]. When these three participants were excluded from the sample, there was still no trend for a correlation between recognition and RT ($r = 0.17$, $p = 0.45$, 95% CI for r = −0.27 to 0.55).

We also examined whether RT priming effects were present in a subgroup of participants who showed no significant behavioral discrimination of words and nonwords on the recognition task ($n = 10$). These participants had the poorest explicit memory, and none of them correctly responded to more than 19/36 trials. Performance of these participants as a group did not exceed chance (defined as 50% correct; mean = 47.2%, SD = 4.3%, $t(9) = -2.02$, $p = 0.074$; *Cohen's d* = 1.35). Despite their poor performance on the recognition task, this group of participants showed robust learning as assessed by the RT measure, with faster RTs for the final position compared to earlier positions [Position effect: $F(2,18) = 11.9$, $p = $

0.002, $\eta^2_p = 0.57$; linear contrast: $F(1,9) = 9.2$, $p = 0.014$, $\eta^2_p = 0.51$; see Figure 3A]. RT effects in this group of participants were not significantly different from those of other participants who performed more accurately on the recognition task [Group x Syllable Position: $F(2,46) = 0.41$, $p = 0.64$; $\eta^2_p = 0.017$; Group x linear position contrast: $F(1,23) = 0.48$, $p = 0.50$, $\eta^2_p = 0.020$; Group x $RT_1 - RT_3$ contrast: $F(1,23) = 0.48$, $p = 0.50$, $\eta^2_p = 0.020$].

One concern is that the absence of a significant correlation between recognition accuracy and RT priming may be driven by a lack of statistical power. Therefore, to increase statistical power we ran an additional analysis that included all participants from both experiments, with the exception of the 3 nonlearners identified in Experiment 2 ($n = 46$). No significant correlation was found between recognition and RT priming ($r = 0.11$, $p = 0.47$, 95% CI for $r = -0.19$ to 0.39).

Item Analysis: We conducted an item analysis to examine whether the relationship between recognition and RT priming differed as a function of participants' performance on the Recognition task. For each participant across the two studies, we categorized each of the six words as Known (comprising items that were correctly recognized on more than 50% of trials on the Recognition task) or Unknown (comprising items that were correctly recognized on 50% or fewer of trials). When the correlation between recognition accuracy and RT priming was restricted to Known words, there was a marginal correlation between Recognition Accuracy and RT priming across all subjects ($n = 49$; $r = 0.25$, $p = 0.081$). However, there were no significant differences in the RT priming effect between Known and Unknown Words (Recognition Classification x Syllable Position ($F(2,96) = 0.92$, $p = 0.40$; $\eta^2_p = 0.019$). Follow-up analyses confirmed that both Known and Unknown words showed significant priming on the RT task (Known words: $t(48) = 5.92$, $p < 0.001$, *Cohen's d = 1.71*; Unknown words: $t(48) = 5.11$, $p < 0.001$, *Cohen's d = 1.47*; Figure 4). Thus, even words that were not successfully recognized in the recognition task elicited robust priming effects.

ERP Results

Recognition Task: We first examined LPC amplitude, including correct and incorrect trials together, as a function of word class. Consistent with our hypothesis, words elicited a significantly larger LPC than nonword foils [$F(1,20) = 8.07$, $p = 0.010$, $\eta^2_p = 0.29$; Figure 5A]. Instruction condition (implicit or explicit) did not significantly impact the LPC word-class effect [Word Class x Instruction Condition: $F(1,20) = 0.45$, $p = 0.51$; $\eta^2_p = 0.022$]. Next, we analyzed LPC amplitude to correct trials alone, in order to examine whether the neural transactions indexed by the LPC are recruited when a correct recognition decision is made. Confirming this hypothesis, correct words elicited a significantly larger LPC than correct nonword foils [Word Class: $F(1,20) = 4.99$, $p = 0.037$; $\eta^2_p = 0.200$]. Instruction condition again did not have a significant effect on LPC amplitude [Word Class x Instruction condition: $F(1,20) = 0.36$, $p = 0.55$; $\eta^2_p = 0.018$].

Target Detection Task: Consistent with our hypothesis, the amplitude of the P300 elicited by target syllables differed as a function of syllable position [Syllable Position: $F(2,46) =$

5.18, $p = 0.010$, $\eta^2_p = 0.18$]. Specifically, initial syllable targets elicited the largest P300, middle syllable targets elicited a moderate P300, and final syllable targets elicited the smallest P300 [linear contrast for Syllable Position: $F(1,23) = 11.9$, $p = 0.002$, $\eta^2_p = 0.34$; Figure 5B]. Instruction condition did not have a significant effect on the P300 syllable position effect [Instruction Condition x Syllable Position: $F(2,46) = 0.94$, $p = 0.40$, $\eta^2_p = 0.039$].

We examined correlations between the behavioral RT priming effect (computed as $RT_{S1} - RT_{S3}$) and the magnitude of the P300 effect (computed as $P300_{S1} - P300_{S3}$) at each of the six electrode groups. Consistent with our predictions, the RT effect correlated significantly with the P300 effect at the right posterior electrode region ($r = 0.42$, $p = 0.037$) and marginally significantly at the midline posterior electrode region ($r = 0.34$, $p = 0.097$). Correlations were positive but did not reach significance at the other four electrode regions (r range = 0.077 – 0.28, p value range = 0.17 – 0.71).

Relationship between Recognition and P300 Priming—Behavioral analyses described previously revealed no significant relationship between recognition accuracy and RT priming on the target detection task. Mirroring this analysis, we examined correlations between recognition accuracy and our ERP measure of priming on the target detection task, the P300 effect. Similar to the null correlation found at the behavioral level, no significant correlations were found (r range = –0.36 – 0.12, all p values > 0.5 for positive r values).

Finally, we investigated whether a similar P300 effect was present in participants who showed no significant behavioral discrimination of words and nonword foils in the recognition task ($n = 10$), as described previously in the behavioral results section. This group showed a significant linear P300 effect following the same pattern as described above [linear contrast for Syllable Position: $F(1,9) = 5.43$, $p = 0.045$, $\eta^2_p = 0.38$; Figure 3B]. The P300 syllable position effect did not significantly differ between high and low performers on the recognition task [Syllable Position x High/Low Recognition Group: $F(2,46) = 0.72$, $p = 0.49$, $\eta^2_p = 0.030$], consistent with results from the correlational analysis.

Discussion

Behavioral and ERP data from the recognition task suggest that accurate recognition judgments are supported by explicit memory. At the behavioral level, accuracy was highest when participants subjectively experienced better or more detailed recollection, and not better than chance when participants reported they were guessing. Thus, according to both the zero-correlation criterion and the guessing criterion, recognition judgments were strongly influenced by explicit knowledge, with no evidence that implicit knowledge contributed. ERP analyses provided additional support for this idea, revealing an enhanced LPC to learned words relative to nonword foils. This effect maps well onto old/new effects observed during explicit memory tasks and interpreted as reflecting recollective processing (Paller & Kutas, 1992; Rugg et al., 1998; Rugg & Curran, 2007; Voss & Paller, 2008). Taken together, these data indicate that statistical learning can produce explicit knowledge, at least in healthy adults.

The target detection task also provided evidence of learning. Behavioral data replicated the pattern observed in Experiment 1, showing that RTs were faster to target syllables in word-final positions and that this measure was uncorrelated with explicit measures. This finding demonstrated that participants were able to make use of the statistical structure of the stimuli during a speeded response task, more quickly executing a response when targets occurred in more predictable positions. P300 data converged with the RT data; amplitude of the P300 scaled linearly with syllable position, showing the largest amplitude to initial syllable targets and the smallest amplitude to final syllable targets. Given that the P300 has been taken to index overall levels of attention and working memory resources needed to process a stimulus (Polich, 2007), this finding indicates that fewer controlled, limited-capacity resources were needed to process targets occurring in more predictable positions. This finding provides evidence against the alternative hypothesis that faster reaction times may be driven by *greater* engagement of controlled, effortful processes. In other words, the behavioral reaction time effect reflects facilitation at a neural level, in which predictable targets are processed more efficiently due to learned cross-syllable patterns. This facilitation could reflect automatic mechanisms such as spreading activation (e.g., Neely, 1991; Collins & Loftus, 1975), in which activation of a given node spreads rapidly and automatically to associated representations. In the context of statistical learning, repeated exposure to the stimulus stream would produce stronger associations between co-occurring syllables. Therefore, activating the representation of a word-initial syllable would increase the activation levels of syllables that typically follow this first syllable, thereby facilitating their processing and reducing both reaction times and P300 amplitudes.

Interestingly, P300 amplitude showed a linear effect as a function of syllable position, whereas significant RT differences emerged only between second and third syllable targets. In our stimuli, the first two syllables predicted the final syllable deterministically, whereas the first syllable predicted the second syllable only probabilistically. The dissociation between the P300 and RT effects suggests that the brain has access to both probabilistic and deterministic information during online processing, but that overt behavioral responses rely primarily upon deterministic cues alone. This reliance upon 100% deterministic cues may occur as a way to minimize errors, wherein a behavioral response is executed only if the learner can establish with certainty that the next stimulus will be a target. According to this idea, although more contextual information is available prior to the onset of second syllable targets relative to first syllable targets, the learner may not prepare his or her response to these targets in advance because their identities cannot be predicted with any certainty. This would lead to faster responding for the final target syllable, and no difference in response times between the first and second target. This idea is consistent with two previous statistical learning studies with stimulus triplets in which both the second and third items were deterministically predicted by the first item (Turk-Browne et al., 2005; Kim et al., 2009), unlike the present stimulus streams. In contrast to the RT pattern we observed, both these studies found a graded RT effect as a function of item position within the triplet. Thus, significant RT differences may emerge only when an item can be uniquely predicted from the preceding context. In contrast, facilitation at the neural level could reflect mechanisms such as spreading activation, which would not necessarily lead to behavioral differences. In

sum, ERPs appear to reflect both probabilistic and deterministic knowledge and may be a more sensitive measure of statistical learning than behavioral measures.

Within the target detection task, the RT priming effect and the P300 effect significantly correlated across participants at the right posterior electrode region. That is, participants who showed a larger decrease in RTs to third syllable targets relative to first syllable targets also showed a larger decrease in P300 amplitude. This finding suggests that RTs and the P300 are both sensitive to a common underlying process, indexing priming effects produced by statistical learning. This finding also demonstrates that, in principle, the signal-to-noise ratio of these data is sufficient to yield significant behavioral-ERP correlations.

In contrast, no significant correlation was found between recognition accuracy and either the RT or P300 effect on the target detection task. In addition, both RT and P300 effects on the target detection task were sensitive to learning even in the absence of explicit recognition. One interpretation of these results is that the mechanisms recruited during the RT task are implicit in nature, operating independently of those that support explicit recognition. Results from the individual item analysis strengthen this possibility. A marginal correlation was found between recognition and RT priming when the analysis was restricted to correctly identified words, suggesting that the RT task may be weakly sensitive to explicit knowledge. However, items that were not correctly identified on the recognition task still elicited robust RT priming effects on the target detection task, and these effects were not significantly different from priming effects observed to correctly identified items. This finding suggests that the target detection task captures learning above and beyond what is accounted for by explicit recognition, and may reflect implicit representations that were accrued in parallel during the statistical learning process. Nonetheless, the interpretation that these two tasks reflect dissociable mechanisms remains somewhat speculative, given the known weaknesses of null correlative evidence.

General Discussion

The strongest conclusion to be drawn from the present study is that statistical learning produces explicit knowledge and that the recognition task appears to largely reflect this explicit knowledge. Until now, the implicit-versus-explicit nature of the memory processes recruited to support performance on this task, by far the most common test of statistical learning, has not been thoroughly examined. We applied methods commonly used in memory studies, namely the remember/know paradigm and the recording of ERPs at test, to critically examine the type of memory processing used to support performance on this task. As a result, we observed that explicit memory significantly influences this measure. Accurate recognition judgments were associated with the experience of either recollection or familiarity, and did not occur when participants reported that they were guessing. In addition, words elicited an enhanced LPC, implicating the involvement of explicit memory retrieval for words relative to nonword foils. It has been previously proposed that performance on the recognition task reflects vague, implicit, intuitive judgments (e.g., Turk-Browne et al., 2005). Our results are inconsistent with this idea, suggesting that participants are aware of the knowledge that they are using to complete this task. Even if implicit

representations are formed during statistical learning, the recognition measure does not appear to be sensitive to them.

Although our results strongly support the idea that statistical learning produces explicit knowledge, it appears that the recognition judgment task may not adequately capture the entire knowledge base that is produced as a result of statistical learning. The target detection task produced faster RTs to final syllables as well as a graded P300 pattern, indicating that learners acquired knowledge that allowed them to more effectively detect syllables in predictable positions. There was also a difference in sensitivity between the two tasks used to assess learning, as demonstrated by the number of participants showing significant effects on each task at the single-subject level (defined as $p < 0.05$, one-tailed test in hypothesized direction). On the recognition task, only 14 participants out of a total of 49 exhibited discrimination that was significantly above chance at the single-subject level (23 trials correct out of 36), whereas 21 participants showed significant RT effects on the target detection task at the individual level (responding significantly faster to third syllable targets relative to first syllable targets). Note that the overall numbers of participants achieving these criteria is somewhat low, as a highly robust level of performance is necessary in order to yield statistical significance at the single-subject level compared to at the group level. Nonetheless, a comparison of these values between the two tasks suggests that the reaction time task was approximately 50% more sensitive to statistical learning effects than the recognition task. Use of this more sensitive test revealed that a subset of participants who would normally be classified as "nonlearners" (exhibiting no explicit recognition for the statistical regularities) did in fact learn the statistical structure of the stimuli implicitly. Although many studies reported higher accuracy on the recognition task than in the present study, it is also the case that relatively low (~60%) or nonsignificant levels of recognition in statistical learning paradigms have been reported previously and are not unusual in the literature (e.g., McNealy et al., 2006; Turk-Browne et al., 2009; Sanders et al., 2002; Arciuli et al., 2014). An RT-based measure such as the target detection task has the additional advantage of showing that statistical learning can actually enhance online performance, improving detection of auditory syllables, rather than merely resulting in above-chance performance on an offline task. Thus, indirect tests such as speeded identification, as traditionally used in implicit learning studies, may provide more sensitive measures of statistical learning, and may also more effectively capture the role or function that statistical learning serves outside the laboratory.

As a secondary point of discussion, performance on the target detection task was uncorrelated with explicit measures and a number of participants who exhibited no explicit recognition for the novel words still showed robust facilitation effects on the speeded task, as demonstrated both by RTs and ERPs. In addition, item analyses revealed that words that were recognized at a level no better than chance still elicited RT priming effects. Although further evidence is needed, these results are consistent with the idea that implicit representations, dissociable from explicit recognition-based knowledge, are produced during statistical learning. Because the best support for a null hypothesis occurs when there is adequate power, we combined participants from both experiments ($n = 46$) to increase power, and still observed a very low correlation between the recognition and target detection

measures ($r = 0.1$). In addition, RT and ERP measures within the target detection task did significantly correlate, indicating that, in principle, the signal-to-noise ratio of the data was adequate to produce significant correlations. The observed dissociations between these two tasks are further supported by two recent behavioral studies of statistical learning that also found dissociations between direct and indirect measures of statistical learning (Kim et al., 2009; Bertels et al. 2013).

Interestingly, we found no effect of instruction on learning performance, either in Experiment 1 or Experiment 2. That is, participants who were informed that the speech stream contained nonsense words and who were asked to discern the identity of these words performed no better on either the recognition or target detection tasks than participants who were simply passively exposed to the underlying regularities. These findings converge with a recent study that compared visual statistical learning under incidental and intentional learning conditions and that also reported no significant difference in the amount of statistical learning as a function of instruction (Arciuli et al., 2014). These authors suggested that presentation rate may be a critical factor in determining whether statistical learning can benefit from explicit instructions to search for underlying regularities. When presentation times are short (e.g., 200 ms duration for each stimulus presented in the Arciuli et al. study), participants may be unable to effectively exploit intentional learning strategies. Our results are generally consistent with this notion, given that each syllable in the present study was separated by less than 300 ms on average. Although the hypothesis of Arciuli and colleagues may well be correct, experimental manipulations designed to test this hypothesis may fail to yield much insight into statistical learning mechanisms that operate outside of the laboratory. These types of manipulation would come at the cost of ecological validity, given the very short duration of typical temporally ordered stimuli in settings outside the laboratory (such as syllables in a stream of speech). Overall, the finding that statistical learning is not readily influenced by the precise instructions given to participants, at least under normal circumstances approximating learning conditions outside the laboratory, supports the general idea that statistical learning proceeds automatically and in the absence of conscious attempts to extract underlying patterns (e.g., Saffran et al., 1997, 1999; Fiser & Aslin, 2001, 2002; Toro, Sinnett & Soto-Faraco, 2005; Turk-Browne et al., 2005, 2009). This important feature of statistical learning allows learners to acquire a large amount of structural information without explicit guidance as to what to learn, making statistical learning a powerful and flexible learning mechanism.

Final implications of these results concern potential improvements that could be made to the design of statistical learning paradigms. First, our results indicate that traditional statistical learning assessments based on rated familiarity reflect explicit memory and should not be taken as measures of implicit learning. Second, our findings suggest that indirect, reaction time-based measures such as those obtained using the present syllable detection task more effectively and selectively measure implicit learning of statistical structure. This type of measure also appears to be a more robust and sensitive measure of statistical learning overall. Future studies of statistical learning might consider incorporating an indirect task in addition to the traditional recognition judgment task in order to more comprehensively assess the degree of learning that has occurred.

Conclusions

Results from the current study suggest that the recognition judgments typically used to assess statistical learning primarily reflect contributions from explicit knowledge. We also demonstrated that statistical learning can be assessed not only using the conventional recognition task, but also using an indirect RT task. The RT task appears to be more sensitive than the recognition task to learning overall and could easily be adopted in future studies of auditory statistical learning. Finally, the finding that recognition is uncorrelated with both behavioral and ERP measures of priming on the RT task suggests the possibility that implicit knowledge is accrued alongside explicit recognition-based knowledge. In contrast to the recognition judgment task, which makes direct reference to prior studied items, the RT task—as a surreptitious test of knowledge accrued during statistical learning—is well-suited to indexing these potential influences from implicit memory.

Acknowledgments

References

Arciuli J, von Koss Torkildsen J, Stevens DJ, Simpson IC. Statistical learning under incidental versus intentional conditions. Frontiers in Psychology. 2014; 5:1–8. [PubMed: 24474945]

Aslin RN, Saffran JR, Newport EL. Computation of conditional probability statistics by 8-month-old infants. Psychological Science. 1998; 9:321–324.

Bates E, Elman J. Learning Rediscovered. Science. 1996; 274:1849. [PubMed: 8984644]

Bertels J, Demoulin C, Franco A, Destrebecqz A. Side Effects of Being Blue: Influence of Sad Mood on Visual Statistical Learning. PloS One. 2013; 8:e59832. [PubMed: 23555797]

Bertels J, Franco A, Destrebecqz A. How implicit is visual statistical learning? Journal of Experimental Psychology: Learning, Memory, and Cognition. 2012; 38:1425–1431.

Collins AM, Loftus EF. A Spreading-Activation Theory of Semantic Processing. Psychological Review. 1975; 82:407–428.

Conway CM, Christiansen MH. Modality-Constrained Statistical Learning of Tactile, Visual and Auditory Sequences. Journal of Experimental Psychology: Learning, Memory & Cognition. 2005; 31:24–39.

Curran T. Higher-order associative learning in amnesia: Evidence from the serial reaction time task. Journal of Cognitive Neuroscience. 1997a; 9:522–533. [PubMed: 23968215]

Curran T. Effects of aging on implicit sequence learning: Accounting for sequence structure and explicit knowledge. Psychological Research. 1997b; 60:24–41. [PubMed: 9225617]

Destrebecqz, A.; Cleeremans, A. Temporal effects in sequence learning. In: Jimenez, L., editor. Attention and implicit learning. Amsterdam: Benjamins; 2003. p. 181-213.

Dienes Z, Berry D. Implicit learning: Below the subjective threshold. Psychonomic Bulletin & Review. 1997; 4:3–23.

Dienes Z, Scott R. Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. Psychological Research. 2005; 69:338–351. [PubMed: 15944859]

Dienes, Z.; Altmann, GTM. Transfer of implicit knowledge across domains: How implicit and how abstract?. In: Berry, D., editor. How implicit is implicit learning!. Oxford, England: Oxford University Press; 1997. p. 107-123.

Dienes Z, Altmann GTM, Kwan L, Goode A. Unconscious knowledge of artificial grammars is applied strategically. Journal of Experimental Psychology: Learning, Memory and Cognition. 1995; 21:1322–1338.

Duncan-Johnson CC, Donchin E. Effects of a priori and sequential probability of stimuli on event-related potential. Psychophysiology. 1977; 14:95.

Duncan-Johnson CC, Donchin E. The P300 component of the event-related brain potential as an index of information processing. Biological Psychology. 1982; 14:1–52. [PubMed: 6809064]

Fiser J, Aslin RN. Unsupervised statistical learning of higher-order spatial structures from visual scenes. Psychological Science. 2001; 12:499–504. [PubMed: 11760138]

Fiser J, Aslin RN. Statistical learning of higher-order temporal structure from visual shape sequences. Journal of Experimental Psychology: Learning, Memory & Cognition. 2002; 28:458–467.

Foerde K, Knowlton BJ, Poldrack RA. Modulation of competing memory systems by distraction. Proceedings of the National Academy of Sciences. 2006; 103:11778–11783.

Franco A, Cleeremans A, Destrebecqz A. Statistical learning of two artificial languages presented successively: How conscious? Frontiers in Psychology. 2011; 2:1–12. [PubMed: 21713130]

Frensch PA, Runger D. Implicit learning. Current Directions in Psychological Science. 2003; 12:13–18.

Friedman D, Johnson R. Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. Microscopy Research and Technique. 2000; 51:6–28. [PubMed: 11002349]

Higham PA. Dissociations of grammaticality and specific similarity effects in artificial grammar learning. Journal of Experimental Psychology: Learning, Memory and Cognition. 1997; 23:1029–1045.

Jacoby LL. A process dissociation framework: separating automatic from intentional uses of memory. Journal of Memory and Language. 1991; 30:513–541.

Jamieson RK, Mewhort DJK. Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity. The Quarterly Journal of Experimental Psychology. 2009a; 62:550–575. [PubMed: 18609412]

Jamieson RK, Mewhort DJK. Applying an exemplar model to the serial reaction-time task: Anticipating from experience. The Quarterly Journal of Experimental Psychology. 2009b; 62:1757–1783. [PubMed: 19219752]

Johnson R, Donchin E. Sequential expectancies and decision-making in a changing environment: an electrophysiological approach. Psychophysiology. 1982; 19:183–200. [PubMed: 7071298]

Kim R, Seitz A, Feenstra H, Shams L. Testing assumptions of statistical learning: Is it long-term and implicit? Neuroscience Letters. 2009; 461:145–149. [PubMed: 19539701]

Kinder A, Shanks DR. Neuropsychological dissociations between priming and recognition: a single-system connectionist account. Psychological Review. 2003; 110:728–744. [PubMed: 14599240]

Kirkham NZ, Slemmer JA, Johnson SP. Visual statistical learning in infancy: evidence for a domain general learning mechanism. Cognition. 2002; 83:B35–B42. [PubMed: 11869728]

Knowlton BJ, Squire LR. The information acquired during artificial grammar learning. Journal of Experimental Psychology: Learning, Memory and Cognition. 1994; 20:79–91.

Knowlton BJ, Squire LR. Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. Journal of Experimental Psychology: Learning, Memory and Cognition. 1996; 22:169–181.

Knowlton BJ, Ramus SJ, Squire LR. Intact artificial grammar learning in amnesia: dissociation of classification learning and explicit memory for specific instances. Psychological Science. 1992; 3:172–179.

Lieberman MD, Chang GY, Chiao J, Bookheimer SY, Knowlton BJ. An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. Journal of Cognitive Neuroscience. 2004; 16:427–438. [PubMed: 15072678]

Manza, L.; Reber, AS. Representing artificial grammar: Transfer across stimulus forms and modalities. In: Berry, D., editor. How implicit is implicit learning?. Oxford: Oxford University Press; 1997.

McNealy K, Mazziotta JC, Dapretto M. Cracking the language code: Neural mechanisms underlying speech parsing. The Journal of Neuroscience. 2006; 26:7629–7639. [PubMed: 16855090]

Meulemans T, Van der Linden M. Associative chunk strength in artificial grammar learning. Journal of Experimental Psychology: Learning, Memory and Cognition. 1997; 23:1007–1028.

Neely, JH. Semantic priming effects in visual word recognition: A selective review of current findings and theories. In: Besner, D.; Humphreys, GW., editors. Basic processes in reading: Visual word recognition. Hillsdale, NJ: Erlbaum; 1991. p. 264-336.

Nissen MJ, Bullemer PT. Attentional requirements for learning: Evidence from performance measures. Cognitive Psychology. 1987; 19:1–32.

Olson IR, Chun MM. Temporal contextual cuing of visual attention. Journal of Experimental Psychology: Learning, Memory and Cognition. 2001; 27:1299–1313.

Paller KA, Kutas M. Brain potentials during memory retrieval provide neurophysiological support for the distinction between conscious recollection and priming. Journal of Cognitive Neuroscience. 1992; 4:375–393. [PubMed: 23968130]

Paller KA, Voss JL, Boehm SG. Validating neural correlates of familiarity. Trends in Cognitive Sciences. 2007; 11:243–250. [PubMed: 17475539]

Perruchet P, Pacteau C. Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge. Journal of Experimental Psychology: General. 1990; 119:264–275.

Perruchet P, Gallego J, Savy I. A critical reappraisal of the evidence for unconscious abstraction of deterministic rules in complex experimental situations. Cognitive Psychology. 1990; 22:493–516.

Polich, J. Overview of P3a and P3b. In: Polich, J., editor. Detection of change: event-related potential and fMRI findings. Boston, MA: Kluwer; 2003. p. 83-98.

Polich J. Updating P300: An integrative theory of P3a and P3b. Clinical Neurophysiology. 2007; 118:2128–2148. [PubMed: 17573239]

Reber A. Implicit learning of artificial grammars. Journal of Verbal Learning and Verbal Behavior. 1967; 6:855–863.

Reber A. Implicit learning of synthetic languages: The role of instructional set. Journal of Experimental Psychology: Human Learning and Memory. 1976; 2:88–94.

Reber PJ, Squire LR. Parallel brain systems for learning with and without awareness. Journal of Learning and Memory. 1994; 1:217–229.

Rugg MD, Curran T. Event-related potentials and recognition memory. Trends in Cognitive Sciences. 2007; 11:251–257. [PubMed: 17481940]

Rugg MD, Mark RE, Walla P, Schloerscheidt AM, Birch CS, Allan K. Dissociation of the neural correlates of implicit and explicit memory. Nature. 1998; 392:595–598. [PubMed: 9560154]

Saffran JR. Constraints on statistical language learning. Journal of Memory and Language. 2002; 47:172–196.

Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. Science. 1996; 274:1926–1928. [PubMed: 8943209]

Saffran JR, Johnson EK, Aslin RN, Newport EL. Statistical learning of tone sequences by human infants and adults. Cognition. 1999; 70:27–52. [PubMed: 10193055]

Saffran JR, Newport EL, Aslin R. Word segmentation: the role of distributional cues. Journal of Memory and Language. 1996b; 35:606–621.

Saffran J, Newport E, Aslin R, Tunick R, Barrueco S. Incidental language learning: Listening (and learning) out of the corner of your ear. Psychological Science. 1997; 8(2):101–105.

Sanchez DJ, Reber PJ. Explicit pre-training instruction does not improve implicit perceptual-motor sequence learning. Cognition. 2013; 126:341–351. [PubMed: 23280147]

Sanders L, Newport EL, Neville HJ. Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. Nature Neuroscience. 2002; 5:700–703.

Scott RB, Dienes Z. The conscious, the unconscious, and familiarity. Journal of Experimental Psychology: Learning, Memory and Cognition. 2008; 34:1254–1288.

Seidenberg M. Language acquisition and use: Learning and applying probabilistic constraints. Science. 1997; 275:1599–1603. [PubMed: 9054348]

Servan-Schreiber E, Anderson JR. Learning artificial grammars with competitive chunking. Journal of Experimental Psychology: Learning, Memory and Cognition. 1990; 16:592–608.

Shanks DR, Johnstone T. Evaluating the relationship between explicit and implicit knowledge in a serial reaction time task. Journal of Experimental Psychology: Learning, Memory and Cognition. 1999; 25:1435–1451.

Shanks DR, Johnstone T. Evaluating the relationship between explicit and implicit knowledge in a sequential reaction time task. Journal of Experimental Psychology: Learning, Memory and Cognition. 1999; 25:1435–1451.

Shanks DR, Perruchet P. Dissociation between priming and recognition in the expression of sequential knowledge. Psychonomic Bulletin & Review. 2002; 9:362–367. [PubMed: 12120801]

Shanks DR, St John MF. Characteristics of dissociable human learning systems. Behavioral and Brain Sciences. 1994; 17:367–447.

Shanks DR, Wilkinson L, Channon S. Relationship between priming and recognition in deterministic and probabilistic sequence learning. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2003; 29:248–261.

Squires KC, Wickens C, Squires NK, Donchin E. Effect of stimulus sequence on waveform of cortical event-related potential. Science. 1976; 193:1142–6. [PubMed: 959831]

Thiessen ED. Domain General Constraints on Statistical Learning. Child Development. 2011; 82:462–470. [PubMed: 21410909]

Toro JM, Sinnett S, Soto-Faraco S. Speech segmentation by statistical learning depends on attention. Cognition. 2005; 97:B25–B34. [PubMed: 16226557]

Tunney RJ, Altmann GTM. Two Modes of Transfer in Artificial Grammar Learning. Journal of Experimental Psychology: Learning, Memory and Cognition. 2001; 27:614–639.

Turk-Browne NB, Junge JA, Scholl BJ. The automaticity of visual statistical learning. Journal of Experimental Psychology: General. 2005; 134:552–564. [PubMed: 16316291]

Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK. Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. Journal of Cognitive Neuroscience. 2009; 21:1934–1945. [PubMed: 18823241]

Vilberg KL, Moosavi RF, Rugg MD. The relationship between electrophysiological correlates of recollection and amount of information retrieved. Brain Research. 2006:161–170. [PubMed: 17027673]

Vokey JR, Brooks LR. Salience of item knowledge in learning artificial grammar. Journal of Experimental Psychology: Learning, Memory and Cognition. 1992; 18:328–344.

Voss JL, Paller KA. An electrophysiological signature of unconscious recognition memory. Nature Neuroscience. 2009; 12:349–355.

Voss JL, Baym CL, Paller KA. Accurate forced-choice recognition without awareness of memory retrieval. Learning and Memory. 2008; 15:454–459. [PubMed: 18519546]

Voss JL, Paller KA. Brain substrates of implicit and explicit memory: The importance of concurrently acquired neural signals of both memory types. Neuropsychologia. 2008; 46:3021–3029. [PubMed: 18691605]

Westerberg CE, Miller BB, Reber PJ, Cohen NJ, Paller KA. Neural correlates of contextual cueing are modulated by an explicit learning strategy. Neuropsychologia. 2011; 49:3439–3447. [PubMed: 21889947]

Willingham DB, Goedert-Eschmann K. The relation between implicit and explicit learning: evidence for parallel development. Psychological Science. 1999; 10:531–534.

Willingham DB, Salidis J, Gabrieli JDE. Direct comparison of neural systems mediating conscious and unconscious skill learning. Journal of Neurophysiology. 2002; 88:1451–1460. [PubMed: 12205165]

Highlights

- Statistical learning was measured through recognition and RT tasks

- Successful recognition associated with stronger recollection and an enhanced LPC

- Statistical learning reduced RTs and P300 amplitude to predictable stimuli

- Recognition and RT effects did not significantly correlate

- Indirect RT-based measures can be sensitive indices of statistical learning

A

B
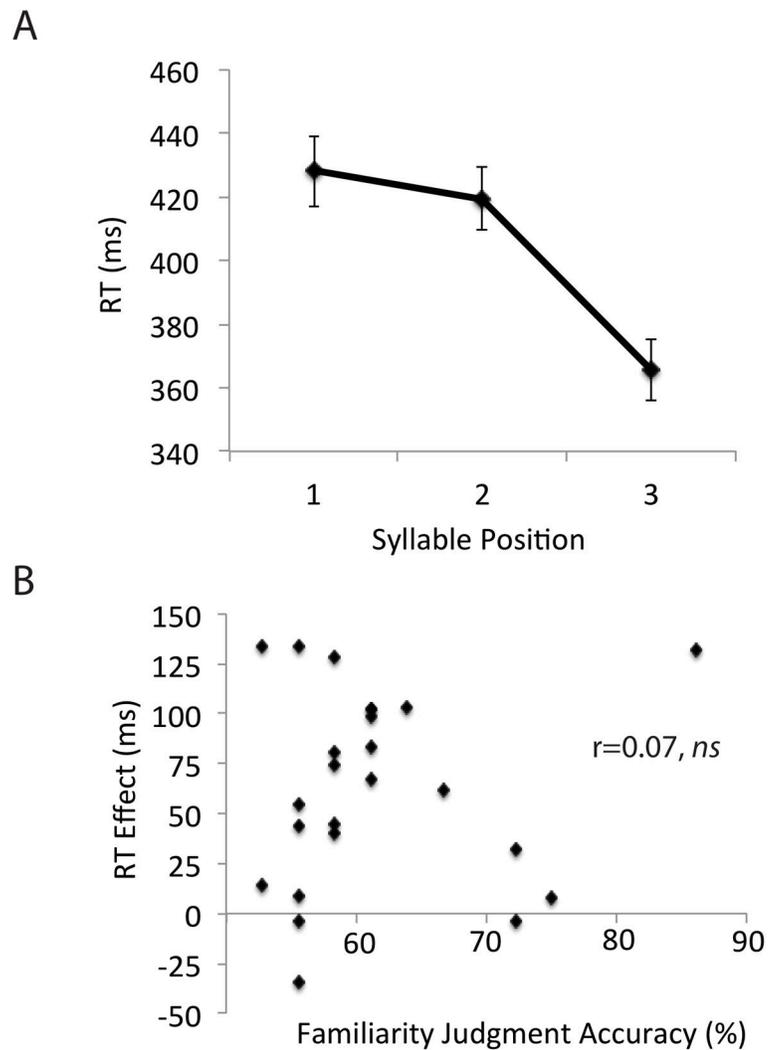
Figure 1.
Behavioral results from Experiment 1. (A) RTs as a function of syllable position in the speeded target detection task. B) Correlation across subjects between accuracy on the recognition task and the magnitude of the RT priming effect on the speeded target detection task, computed by subtracting each subject's median RT to word final targets from his or her median RT to word initial targets. Error bars represent SEM.
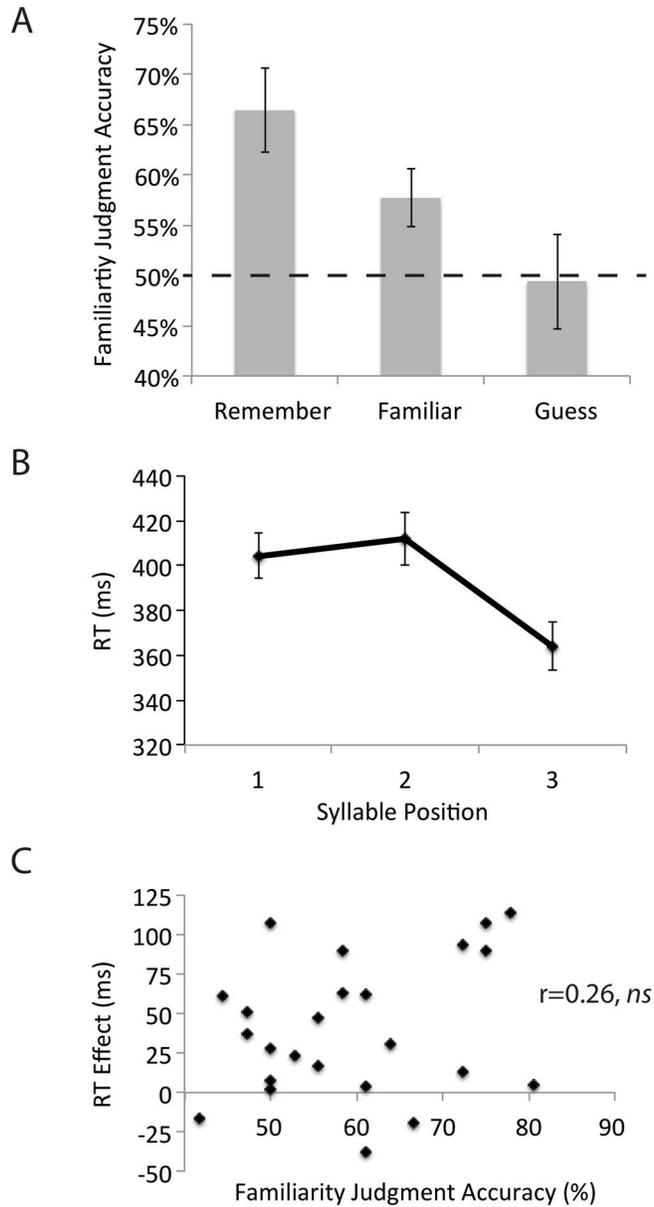
Figure 2.
Behavioral results from Experiment 2. (A) Accuracy on the recognition task as a function of meta-memory judgment. Chance performance on this task is 50%. (B) RTs as a function of syllable position in the speeded target detection task. (C) Correlation across subjects between accuracy on the recognition task and the magnitude of the RT priming effect on the speeded target detection task. Error bars represent SEM.

A
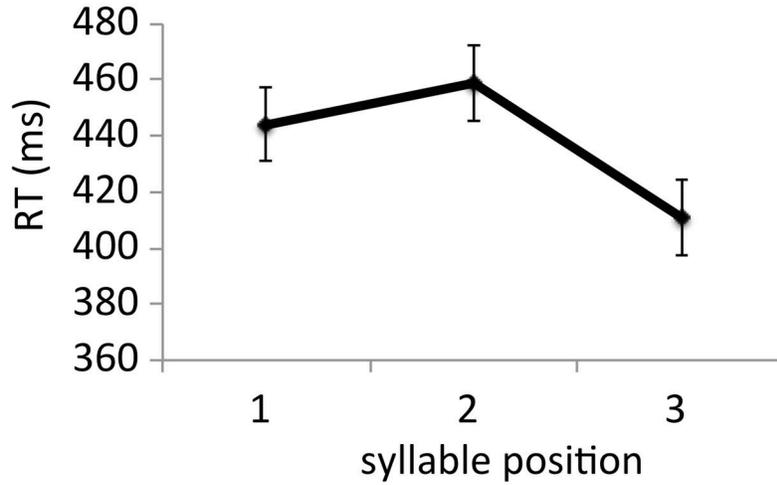
B
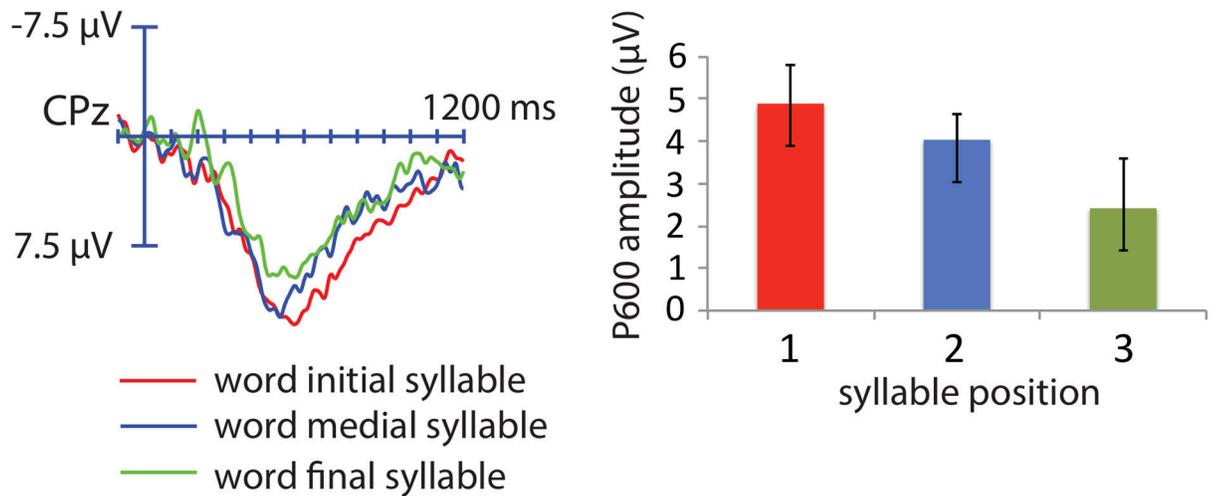
Figure 3.
Data from the speeded target detection task in a subset of participants who scored at chance or below on the recognition task (19 or fewer items correct out of 36; $n = 10$). (A) RTs as a function of syllable position. (B) ERPs timelocked to targets as a function of syllable position. The bar graph shows ERP amplitude to each of the three types of targets across all electrodes. No significant differences, either behavioral or ERP, were found in this subgroup of participants compared to participants who achieved higher accuracies on the familiarity task. ERPs are filtered at 30 Hz for presentation purposes. Error bars represent SEM.
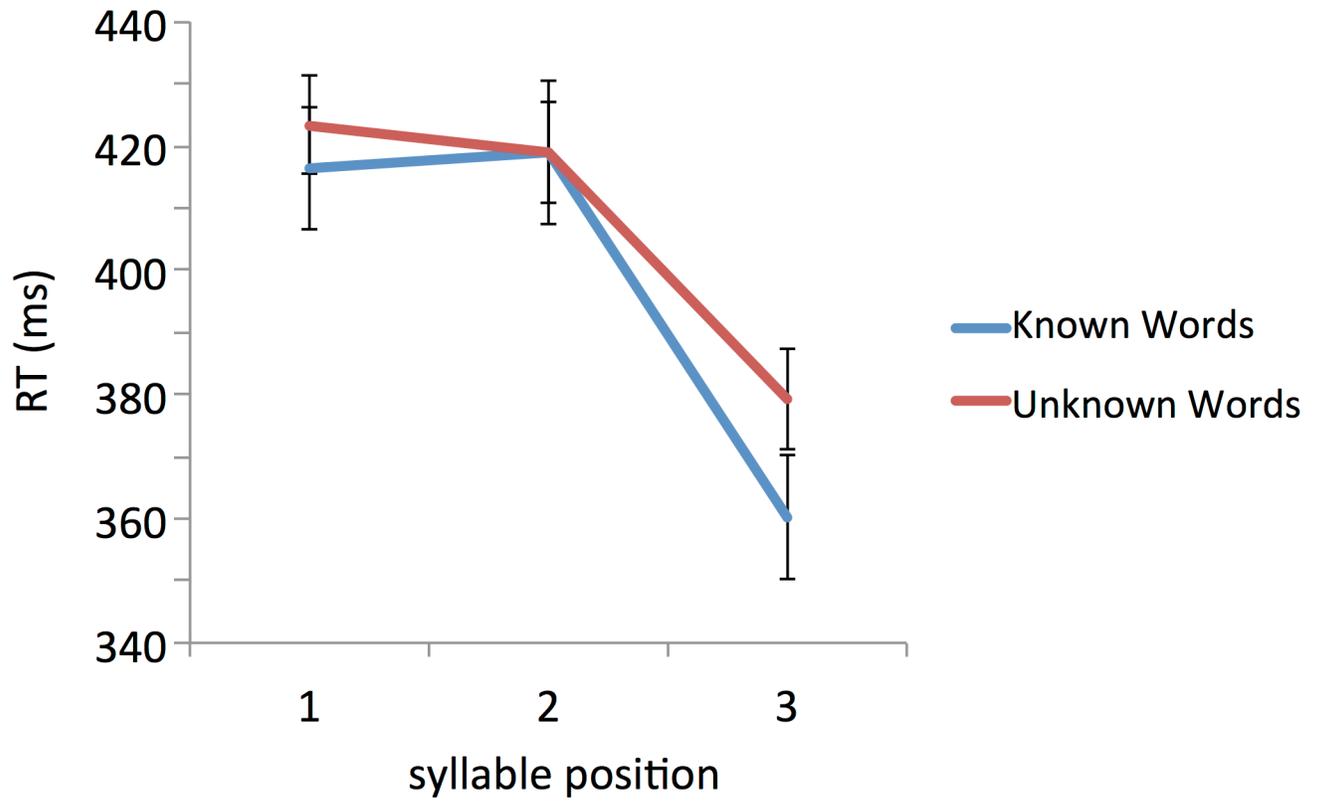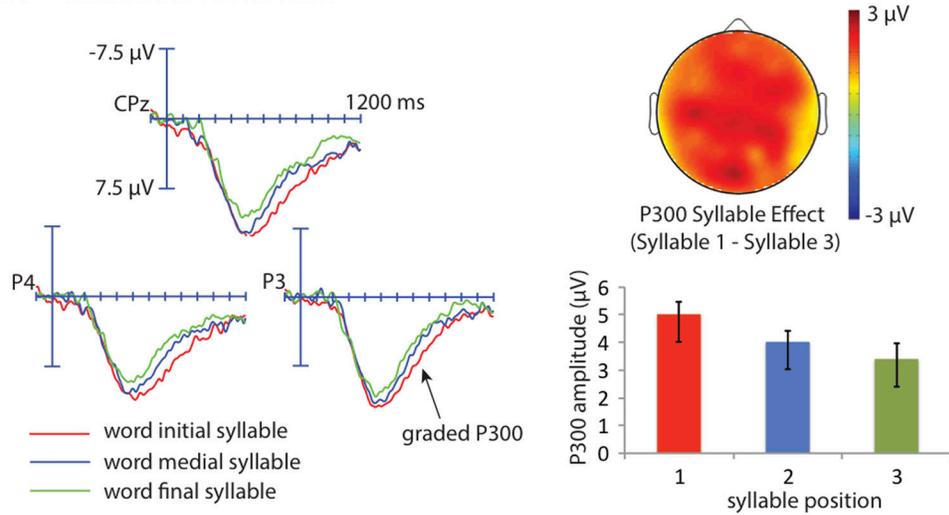
Figure 4.
Results of item analysis examining RTs on the target detection task as a function of performance on the familiarity task. RTs are plotted as a function of syllable position in the speeded target detection task, shown separately for Known words (defined as words that were correctly recognized on more than 50% of trials in the familiarity task) and Unknown words (defined as words that were correctly recognized on 50% or fewer of trials). No significant RT priming differences were found between these two types of items. Error bars represent SEM.
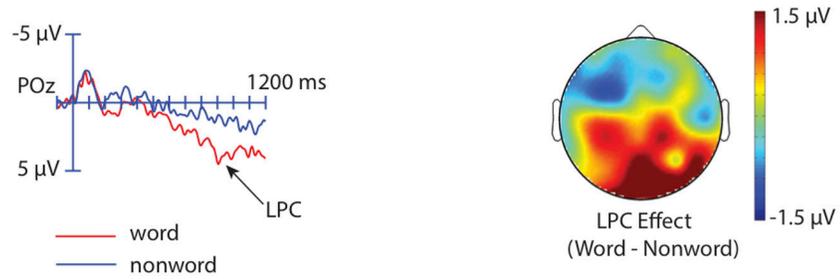
Figure 5.
ERP results from Experiment 2. (A). ERPs timelocked to words and nonword foils in the recognition task. Both correct and incorrect trials are included in these averages. Scalp topography of the late positive component effect, averaged from 700 to 1000 ms, is shown on the right. (B) ERPs timelocked to targets occurring in word initial, word middle, and word final positions in the speeded target detection task. Only correctly detected targets are included in these averages. Scalp topography on the P300 syllable effect, averaged from 400 to 800 ms, is shown on the right. The bar graph displays mean ERP amplitudes across all electrodes as a function of target position. ERPs are filtered at 30 Hz for presentation purposes. Error bars represent SEM.