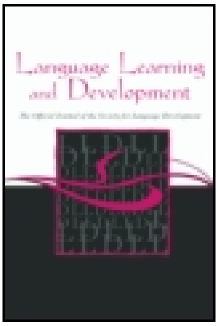

Language Learning and Development
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hlld20

Probabilistically Cued Patterns Trump Perfect Cues in Statistical Language Learning

Jill Lany [a] & Rebecca L. Gómez [b]
[a] Department of Psychology, University of Notre, Dame
[b] Department of Psychology, The University of Arizona
Version of record first published: 30 Nov 2012.

Language Learning and Development, 00: 1–22, 2012

ISSN: 1547-5441 print / 1547-3341 online
DOI: 10.1080/15475441.2012.685826

Psychology Press
Taylor & Francis Group

Probabilistically Cued Patterns Trump Perfect Cues in Statistical Language Learning

Jill Lany
Department of Psychology, University of Notre Dame

Rebecca L. Gómez
Department of Psychology, The University of Arizona

Probabilistically cued co-occurrence relationships between word categories are common in natural languages but difficult to acquire. For example, in English, determiner-noun and auxiliary-verb dependencies both involve co-occurrence relationships but determiner-noun relationships are more reliably marked by correlated distributional and phonological cues and appear to be learned more readily. We tested whether experience with co-occurrence relationships that are more reliable promotes learning those that are less reliable using an artificial language paradigm. Prior experience with deterministically cued contingencies did not promote learning of less reliably cued structure, nor did prior experience with relationships instantiated in the same vocabulary. In contrast, prior experience with probabilistically cued co-occurrence relationships instantiated in different vocabulary did enhance learning. Thus, experience with co-occurrence relationships sharing underlying structure but not vocabulary may be an important factor in learning grammatical patterns. Furthermore, experience with probabilistically cued co-occurrence relationships, despite their difficultly for naïve learners, lays an important foundation for learning novel probabilistic structure.

Natural languages contain co-occurrence relationships between word categories that correspond with important grammatical patterns. For example, in English, functional-elements (e.g., determiners such as *a* and *the* and auxiliary verbs such as *is* and *was*) tend to precede open-class elements that convey semantic information (e.g., nouns and verbs such as *baby* and *drinking*). Thus, nouns and verbs can be distinguished from each other by their distributional properties, or by the sentence contexts in which they occur (Mintz, Newport, & Bever, 2002). They also differ on a host of phonologic properties: For example, nouns tend to have simple consonant onsets, strong-weak stress patterns, and end in the diminutive inflection "y," while verbs tend to begin with consonant clusters, have weak-strong stress patterns, and end in the progressive inflection "ing" (Christiansen, Onnis, & Hockema, 2009; Kelly, 1992; Monaghan, Chater, & Christiansen, 2005). Infants and adults successfully group words into different categories and learn their co-occurrence relationships when they have both distinct phonological properties and

Correspondence should be addressed to Jill Lany, Psychology Department, University of Notre Dame, Haggar Hall, 118-C, Notre Dame, IN 46566. E-mail: jlany@nd.edu

distinct distributional properties (Frigo & McDonald, 1998; Gerken, Wilson, & Lewis, 2005; Gómez & Lakusta, 2004).

While natural languages incorporate such correlated distributional and phonological cues to syntactic categories (Farmer, Monaghan, & Christiansen, 2006; Monaghan et al., 2005; Monaghan, Christiansen, & Chater, 2007), there is substantial variability in the consistency with which such cues are manifested. This variability influences learning, as children more readily acquire co-occurrence relationships in languages in which these structures are reliably cued, such as Italian, than in languages such as English in which structures are less reliably cued (Devescovi et al., 2005; Pizzuto & Caselli, 1992). In studies using artificial language materials, adults (Braine, 1987) and 17-month-old infants (Gerken et al., 2005) successfully learn co-occurrence relationships only when at least 50% of the words within each category have distinctive phonological properties (e.g., when 50% of nouns and verbs contain category-specific phonology). These findings provide converging evidence that sensitivity to category-level co-occurrence relationships begins to break down when they are less reliably marked by distributional and phonological cues, and raise the question of how this factor impacts learning. We lay down a series of hypotheses addressing the question below.

One possibility (Hypothesis 1) is that experience with more reliably cued structures of language may play a role in successful acquisition of similar but less reliably cued patterns. For example, determiner-noun and auxiliary-verb co-occurrence relationships have similar underlying structure (i.e., the reliable co-occurrence of functors that primarily serve a grammatical role with open-class words that convey semantic content). However, a corpus analysis of child-directed speech suggests that nouns are much more reliably cued by inflectional morphology than verbs: Nouns occur with a determiner and/or plural or diminutive ending 82% of the time, while verbs occur with an auxiliary and/or tense marker only 21% of the time (see Lany et al., 2007). Children also appear to learn these properties of nouns more readily than verbs, using newly taught nouns in novel grammatical structures and with novel grammatical morphology in their second year, but failing to show similar generalization for verbs (Tomasello & Olguin, 1993). Despite the fact that determiner-noun and auxiliary-verb structures have minimal vocabulary overlap, learners may nonetheless benefit from their underlying similarity if they are more likely to detect the less reliably marked structure after learning the more reliably marked one. If so, we could also ask how much reliability is necessary for facilitation to occur between one learning instance and the next.

In addition, experience with multiple co-occurrence structures versus just one type might be an important factor in learning abstract structure (Hypothesis 2). Indeed, exposure to variable or diverse instances of a pattern often promotes learning abstract structure and subsequent generalization (e.g., Fried & Holyoak, 1984; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990). Gentner and colleagues have suggested that the process of comparing different exemplars allows learners to perceive abstract similarities between analogous elements within a pattern (e.g., Gentner & Markman, 1997; Gentner & Medina, 1998; Gentner & Namy, 1999). Building on such findings, the current experiment tested how the acquisition of grammatical co-occurrence relationships is affected by prior experience. In particular, given that greater abstraction often results from encountering exemplars with different or more varied surface characteristics (e.g., Gentner & Markman, 1997; Osherson et al., 1990), it is possible that experience with both determiner-noun and auxiliary-verb co-occurrence relationships results in better learning of the abstract structure of those instances than experience with either of these structures alone.

Hypotheses 1 and 2 are orthogonal, and thus if learners do benefit from exposure to variable surface features, we can ask whether experience with *more reliably-cued* structure promotes learning *less reliably cued* structure, which should presumably be more difficult to learn. Smith and colleagues have found that learning a pattern tunes attention to the relevant properties of novel input, thus accelerating and strengthening subsequent learning (Colunga & Smith, 2003, 2005; Jones & Smith, 2002; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002). Thus, experience with reliable co-occurrence relationships between functors and phonological features might attune learners to those elements in similar structures that are not cued reliably enough to capture attention on their own, thereby promoting learning (Hypothesis 1). In addition, because exposure to diverse instances with common structure can promote learning, prior experience with less reliably cued structure may also support learning (Hypothesis 2).

There are two other mechanisms by which learners might capitalize on prior experience when exposed to a pattern with similar underlying structure. First, prior experience with a different pattern may reduce processing demands for the learner despite the fact that the items instantiating the shared pattern differ in their perceptual properties or surface structure. On this account, prior experience with a pattern containing similar underlying structure would facilitate learning *in spite of* differences in the words themselves rather than *because of* these differences (as proposed by H2), and thus similar or even greater benefits should result from giving participants extra experience with the same structure (Hypothesis 3). Another possibility (Hypothesis 4) is that the structural properties of language stimuli are inherently represented in terms of abstract rules, which would permit free generalization to novel exemplars regardless of input properties (e.g., Marcus, Vijayan, Bandi Rao, & Vishton, 1999).

To test these questions, we varied prior experience with an artificial language for five groups of adult participants before they were exposed to probabilistically cued co-occurrence relationships: Specifically, the language contained two content-word-like categories, and only 67% of words in each category were cued by distinctive phonology (see Table 2). Infants require much higher levels of cueing to abstract word categories in a similar artificial language (Gomez & Lakusta, 2004), and pilot research with the artificial language used in the current study indicated that adults fail to abstract when only 67% of words contain phonological cues. In the present study, one group was given no prior experience—their performance served as a baseline measure of learning (67% Naïve control condition). A second group was given prior exposure to the same 67% cued contingencies to assess whether additional experience with specific strings in the same vocabulary, rather than surface variability, facilitates learning probabilistically cued co-occurrence relationships (67%/Same Language condition). Finally, three groups were given prior experience with different co-occurrence relationships before training on the 67% cued language. In these conditions, the artificial language in the pre-exposure phase also contained co-occurrence relationships between word categories but differed in the degree to which the relationships were marked by correlated cues, with 67%, 83%, or 100% of words from different categories cued by distinctive phonology (the 67/67%, 83/67%, and 100/67% conditions).

The manipulation of prior experience across multiple conditions allowed us to test the conditions under which experience affects subsequent learning. If experience with variable surface characteristics acts to highlight the abstract co-occurrence structure (Hypothesis 2), then we should see greater learning for groups given exposure to a different language than to the same language. If benefits arise instead from reduced processing demands (e.g., Hypothesis 3: increased facility with the computations that are critical for learning), the participants given additional exposure to the same language should benefit. However, if learners do benefit from variability in surface structure, we can test how the reliability of the cues marking the co-occurrence relationships affects learning. One possibility, consistent with Hypothesis 1, is that experience with more reliably cued contingencies (i.e., a 100% Cued pattern) will result in robust learning that will better promote learning of less reliably cued contingencies (i.e., a 67% Cued pattern), with the benefits decreasing with decreases in the cue-strength of the previously acquired structure. However, it is also possible that a closer match in the underlying structure is important for facilitating subsequent learning, such that the benefits from prior experience decrease as the number of words cued by distinctive features (67%, 83%, or 100%) increases (Hypothesis 2). Equally high performance across conditions would be consistent with Hypothesis 4.

We staged exposure to the two patterns because it allowed us to obtain a measure of learning of the initial pattern before participants received additional training. If we had used simultaneous presentation it would be unclear how the two patterns affect one another, obscuring the directional effects. We chose to test adult participants because there is ample evidence that changes in learning as a function of prior experience can be observed both in infants and adults in the acquisition of co-occurrence relationships such as the ones we are testing (Lany et al., 2007; Lany & Gómez, 2008). Furthermore, with adult participants we can present multiple types of test items to the same participants to obtain nuanced information about sensitivity to the structure in both phases of learning, which is not possible with infant-testing methods. Thus, while it will ultimately be important to investigate how the process is similar or different in infants, initial testing with adults can help shed light on important questions about the mechanisms by which prior experience affects learning grammatical patterns.

We chose to test these questions using artificial language materials. While these materials were substantially less complex than related structures in natural language, this approach allowed us to achieve precise control over the cues presented to learners and the kinds of prior experiences they were afforded. In addition, previous studies testing infants' ability to learn co-occurrence structure suggests that it connects with other important properties of grammatical categories. For example, infants readily integrate information about the distributional and phonological cues marking word categories with their semantic properties (Lany & Saffran, 2010). Moreover, infants who are better able to capitalize on distributional and phonological cues in word-learning tasks also have higher levels of native-language proficiency (Lany, 2012; Lany & Saffran, 2011). Thus, there is evidence that testing the learning processes underlying sensitivity to this particular type of artificial language structure can shed light on mechanisms supporting natural language acquisition.

METHOD

Participants

Participants were 210 monolingual English-speaking students at the University of Arizona free of hearing loss or a language disorder. An additional 30 students participated, but their data were excluded for giving grammaticality judgments of all "yes" or all "no" (N = 28, see also Procedure section), or because of equipment failure (N = 2). Participants were randomly assigned to one of the five familiarization conditions listed in Table 1 (N = 40 in the Naïve control condition,

TABLE 1
Language Exposure in the 5 Familiarization Conditions

Condition	*Phase 1*	*Phase 2*
Naive		67% Cued (Version B)
100/67% Cued	100% Cued (Version A)	67% Cued (Version B)
83/67% Cued	83% Cued (Version A)	67% Cued (Version B)
67/67% Cued	67% Cued (Version A)	67% Cued (Version B)
67%/Same Language	67% Cued (Version B)	67% Cued (Version B)

Note: In counterbalanced conditions participants were exposed to Version B in Phase 1 and Version A in Phase 2.

N = 44 in the Same Language condition, and N = 42 in each of the 67/67%, 83/67%, and 100/67% conditions). Participants received course credit for their participation.

Materials

Familiarization. The familiarization materials consisted of an *aX bY* language adapted from a previous study investigating adults' ability to learn co-occurrence relationships between word categories (Lany et al., 2007). The language consisted of nonsense words belonging to the categories *a*, *b*, *X*, and *Y*. Words were combined into strings of the form *aX* and *bY,* or, in a counterbalanced condition, *aY* and *bX*. This structure is similar to determiner-noun and auxiliary-verb co-occurrence relationships in English. To test how prior experience influences learning probabilistic co-occurrence relationships, we constructed 2 versions (A and B) of the *aX bY* language. The versions differed only in the words used to instantiate the pattern (see Tables 2a and b). In each version, there were two each of the monosyllabic *a*- and *b*-words, and six each of the *X*- and *Y*-words. The *X*s and *Y*s were disyllabic, but they were distinguished from each other by a phonological cue. In Version A, *X*s ended in the syllable "it" (e.g., *feegit*, *lepit*), and *Y*s ended in the syllable "oo" (e.g., *juhnoo*, *tamoo*), while in Version B, *X*s ended in "ul" and *Y*s ended in

TABLE 2a
Version A Language Materials

a	*b*	*X*	*Y*
		Cued	
ong	erd	bivul	nusee
rud	vot	choopul	lemee
		habbul	sufee
		jerul	vaymee
		pogul	rafee
		vummul	durpee
		Uncued	
		pefto	safon
		bowda	veelay

TABLE 2b
Version B Language Materials

a	*b*	*X*		*Y*
			Cued	
ush	alt	kirit		juhnoo
dak	pel	feegit		tamoo
		soolit		feenoo
		yohvit		zinoo
		zamit		deechoo
		lepit		wifoo
			Uncued	
		jeeloff		skiger
		shaleb		jula

Note: Tables 2A and 2B depict the language materials for Version A and Version B. For each version, the specific *a*, *b*, *X*, and *Y* elements listed were combined to form *aX* and *bY* strings in G1, and in G2 they were combined to form *aY* and *bX* strings. The 83% Cued participants heard the first row of uncued *X*s and *Y*s in place of 2 of the cued *X* and *Y* elements, and the 67% Cued participants heard all 4 Uncued *X*s and *Y*s in place of 4 of the Cued *X* and *Y* elements.

"ee." Thus, each version of the *aX bY* language contained correlated cues distinguishing words from the *X* and *Y* categories: 1) *X*s and *Y*s had distinct distributional properties (i.e., *X*s and *Y*s occurred in different contexts depending on whether they were preceded by an *a*- or a *b*-word), and 2) and they also had distinct phonological properties.

Within each version, there were also two different grammars such that in Grammar 1 strings took the form *aX* and *bY*, and in Grammar 2 they took the form *aY* and *bX*. This manipulation served to rule out effects specific to particular word or feature combinations. For ease of reference, we use the notation "*aX bY*" to describe the materials and structure of this language more generally, but it should be noted that the opposite pairings held in G2.

Studies employing variants of this artificial language have revealed that when the *X*s and *Y*s differ only in their distributional properties, learners demonstrate memory for strings they were trained on, and also learn positional information such as whether a word occurs in string-initial or string-final position. However, under these conditions they do not learn abstract co-occurrence relationships, as reflected in their failure to generalize to unheard strings (Smith, 1969). In contrast, when words from the *X* and *Y* categories have distinct phonological properties in addition to distinct distributional ones, infants and adults do learn the abstract co-occurrence relationships (Frigo & McDonald, 1998; Gerken et al., 2005; Gómez & Lakusta, 2004). The joint presence of distributional and phonological cues appears to facilitate learning by reducing computational and memory demands on learners. Rather than having to remember each individual *aX* or *bY* combination, learners can track the simpler co-occurrence relationships between *a*s and one phonological feature, and between *b*s and a different phonological feature. The relationships between *a*s and *b*s and distinctive phonological features are referred to as marker-feature relationships because the *a*s and *b*s resemble categories that mark a grammatical function (as opposed to conveying semantic information). Learners sensitive to these marker-feature relationships can generalize to unattested strings in which *a*-and *b*-elements are paired with novel *X*- and *Y*-elements, as long as they contain the distinctive phonological feature.

Upon learning that *a*s and *b*s predict words with different phonological endings, learners are also able to incorporate novel *X* and *Y* instances into the paradigm even when they lack these endings based on the presence of an *a*- or *b*-element alone (Frigo & McDonald, 1998). Thus, sensitivity to the marker-feature relationships is an important component of learning the co-occurrence relationships between word categories *per se*, i.e., the higher-level regularity in which *a*s are followed by one set of words, and *b*s are followed by a different set. Such learning would be evidenced by the fact that, upon hearing the *aX* string *ong pefto* from Table 2A, learners generalize to *rud pefto,* while rejecting the ungrammatical *erd pefto* or vot *pefto*, even though *pefto* is not marked by a distinctive ending cueing its category membership such as "it" or "oo" (see Tables 2a and b). This level of sensitivity is more abstract in that it reflects generalizing beyond the concrete marker-feature relationships.

Cue-probability manipulation. Both Versions A and B of the *aX bY* language varied in the number of *X*s and *Y*s containing distinctive phonological features. In the 100% Cued language, all of the *X*s and *Y*s contained the distinctive phonological feature (i.e., in Version A, 6/6 *X*s ended in "it" and 6/6 *Y*s ended in "oo"). In the 83% Cued language, 5/6 of the *X*s and *Y*s contained the cues, and in the 67% Cued language 4/6 *X*s and *Y*s were cued. In all cases, the *X*s and *Y*s lacking the distinctive endings were disyllabic, but the second syllable did not contain a phonological cue to category membership (see Tables 2a and b).

Combining each of the two *a*s with each of the six *X*s yielded 12 *aX* strings, and combining the two *b*s with the six *Y*s yielded 12 *bY* strings, resulting in a total of 24 grammatical strings. However, in each language some of these strings were withheld from familiarization to assess generalization at test. In the 100% condition, the four withheld strings all contained phonological cues. In the 83% condition, one *aX* and one *bY* string with phonological cues were withheld, and one *aX* and one *bY* string lacking phonological cues were withheld (for a total of four withheld strings). In the 67% Cued language, two strings of each type were withheld (for a total of eight withheld strings). Tables 3a and b contain the Generalization +Feature and Generalization –Feature strings for the different language versions.

The language materials were spoken by a female in an animated voice, and were recorded and digitized for editing. The same talker recorded materials for Versions A and B of the language. The same tokens of each word were used in both grammars (e.g., in Version A the same token of *ong* was combined with *X*s in G1 and with *Y*s in G2), and thus the two grammars of each version differed only in the way that words were combined into strings. Strings were approximately 1.7 s in duration, and were separated by 1 s of silence when presented during familiarization. Words within a string were separated by 100 ms of silence.

Test. Test materials consisted of both grammatical and ungrammatical strings (see Tables 3a and b). There were four kinds of grammatical strings crossing whether a string had been presented (or heard) during familiarization, and whether the *X*- or *Y*-word in the string was marked by a distinctive ending, or feature. First, there were **Familiar +Feature** strings, which had been heard by participants during familiarization, and in which the *X*s and *Y*s were marked by the distinctive endings (e.g., *rud choopul* in Version A, G1). The **Familiar –Feature** strings had also been heard during familiarization, but the *X*s and *Y*s in these strings lacked the distinctive endings (e.g., *ong pefto* in Version A G1). For each version of the language, generalization strings were *aX* or *bY* combinations that were not presented during familiarization (e.g., ong *vummul*),

TABLE 3A
Version A Test Strings

100% Cued Test Strings	
Familiar +Feature	*Generalization +Feature*
Grammatical Test Strings	
ong vummul	ong choopul
rud pogul	rud bivul
erd vaymee	erd nusee
vot durpee	vot lemee
Ungrammatical Test Strings	
ong vayme	ong nusee
rud durpee	rud lemee
erd vummul	erd choopul
vot pogul	vot bivul

Version A 83% Cued Test Strings			
Familiar +Feature	Familiar −Feature	Generalization +Feature	Generalization −Feature
Grammatical Test Strings			
rud choopul	ong pefto	ong vummul	rud pefto
vot lemee	erd safon	erd rafee	vot safon
Ungrammatical Test Strings			
vot choopul	ong safon	ong rafee	rud safon
rud lemee	erd pefto	erd vummul	vot pefto

Version A 67% Cued Test Strings			
Familiar +Feature	Familiar −Feature	Generalization +Feature	Generalization −Feature
Grammatical Test Strings			
ong bivul	ong pefto	ong pogul	ong bowda
rud choopul	rud bowda	rud vummul	rud pefto
erd sufee	erd veelay	erd vaymee	erd safon
vot rafee	vot safon	vot nusee	vot veelay
Ungrammatical Test Strings			
ong sufee	ong veelay	ong vaymee	ong safon
rud rafee	rud safon	rud nusee	rud veelay
erd bivul	erd pefto	erd pogul	erd bowda
vot choopul	vot bowda	vot vummul	vot pefto

but did contain an X or Y that had been combined with a different marker in a string that was presented (e.g., the string *rud vummul* had been heard). The **Generalization +Feature** strings were grammatical strings that had been withheld from familiarization and that were marked by the distinctive phonological endings. The **Generalization –Feature** strings were also grammatical strings that been withheld from familiarization, but the Xs and Ys lacked distinctive endings. Because strings that were grammatical in G1 were ungrammatical to participants familiarized to G2, the ungrammatical strings were simply the corresponding string from the other grammar. For

TABLE 3B
Version B Test Strings

Version B 100% Cued Test Strings	
Familiar +Feature	*Generalization +Feature*
Grammatical Test Strings	
ush sulit	ush lepit
dak zamit	dak kirit
alt feenoo	alt juhnoo
pel zinoo	pel wifoo
Ungrammatical Test Strings	
ush feenoo	ush juhnoo
dak zinoo	dak wifoo
alt sulit	alt lepit
pel zamit	pel kirit

Version B 83% Cued Test Strings			
Familiar +Feature	Familiar −Feature	Generalization +Feature	Generalization −Feature
Grammatical Test Strings			
dak feegit	ush geeloff	ush zamit	dak geeloff
pel tamoo	alt skiger	alt juhnoo	pel skiger
Ungrammatical Test Strings			
dak wifoo	ush skiger	ush juhnoo	dak skiger
pel feegit	alt geeloff	alt zamit	pel geeloff

Version B 67% Cued Test Strings			
Familiar +Feature	Familiar –Feature	Generalization +Feature	Generalization –Feature
Grammatical Test Strings			
ush yohvit	ush geeloff	ush feegit	ush shaleb
dak zamit	dak shaleb	dak kirit	dak geeloff
alt zinoo	alt jula	alt tamoo	alt skiger
pel deechoo	pel skiger	pel wifoo	pel jula
Ungrammatical Test Strings			
ush zinoo	ush jula	ush tamoo	ush skiger
dak deechoo	dak skiger	dak wifoo	dak jula
alt yohvit	alt geeloff	alt feegit	alt shaleb
pel zamit	pel shaleb	pel kirit	pel geeloff

Note: Table 2 depicts the full set of test strings for each of Cue Levels in Versions A and B. The strings listed as Grammatical were in fact grammatical for G1, and the strings listed as Ungrammatical served as the Grammatical test strings for participants exposed to Grammar 2. The test strings listed as Generalization (+ or –Feature) were those withheld from familiarization.

example, *rud choopul* was a grammatical Familiar +Feature string in G1 of Version A, and its corresponding foil, *vot choopul*, was ungrammatical, while the opposite was true for participants exposed to G2.

Participants could discriminate between Familiar strings (+Feature and –Feature) and ungrammatical ones entirely on the basis of familiarity or memory for which strings had been heard versus those that had not been heard. However, participants could only discriminate between grammatical and ungrammatical Generalization strings on the basis of having learned the language's co-occurrence relationships. In the case of Generalization +Feature strings, successful discrimination could be accomplished by recalling the marker-feature co-occurrence relationships. For the Generalization –Feature strings, in which the *X*s and *Y*s lacked the phonological features cueing category membership, participants could only discriminate the grammatical strings from the ungrammatical ones if they had abstracted the higher dimension *aX bY* co-occurrence restrictions that are not dependent on a feature being present.

In the 100% and 83% Cued conditions, there were 16 unique test strings, half of which were grammatical and half ungrammatical. The 100% Cued language contained only strings with features, and thus the test materials for this language consisted of four Familiar +Feature strings and four Generalization +Feature strings, as well as their ungrammatical foils. The test for the 83% Cued condition consisted of two each of the 4 grammatical types: Familiar +Feature, Familiar –Feature, Generalization +Feature strings, and Generalization –Feature strings, and their ungrammatical foils (see Table 3). The test for the 67% Cued language contained four strings of each kind for a total of 32 unique test strings. Test strings had the same acoustic characteristics as the familiarization strings.

Design and Procedure

There were five conditions (see Table 1), each of which consisted of exposure to 67% Cued *aX bY* co-occurrence relationships as depicted in the column labeled "Phase 2" in Table 1. Critically, the groups differed in their prior experience with the *aX bY* language, as can be seen in the column labeled "Phase 1." Participants in the Naïve control condition were trained and tested on a 67% Cued *aX bY* language, with no prior experience, and with version and grammar counterbalanced across participants. The remaining four conditions consisted of two consecutive train-test phases. In the 67%/Same Language condition, participants were trained and tested on the same 67% Cued language in both phases. In the remaining three conditions, participants were given prior experience with a different 100%, 83%, or 67% Cued version of the *aX bY* language before being trained and tested on a 67% Cued language in the second phase. For instance, in the 100/67% Cued condition, participants were first exposed to Version A of the 100% Cued language and then to Version B of the 67% Cued language (or, to Version B of the 100% Cued language and then Version A of 67% Cued language) and so on for the 83/67% and the 67/67% conditions.

Participants were individually tested on computers. At the start of the experiment, participants in all conditions were instructed that they would listen to a nonsense language, and they should pay close attention because they would later be tested on what they had learned. They then listened to 18 randomized blocks of the familiarization strings over headphones. This phase took about 18 minutes. Participants then began the test phase, in which they were instructed that the strings in their nonsense language followed a pattern. They were told to listen to a series of strings and make a judgment as to whether each string followed the same pattern as in the familiarization phase. They were also told that half of the strings followed the pattern while the other half did not, and that half of their answers should thus be "yes" and half should be "no." Following the instructions, participants were presented with one randomized block of the test strings. The instructions were then repeated and a second block of test trials was presented. The familiarization and test materials were presented using Superlab Pro software. Participants made their responses at test by pressing the "Y" and "N" keys on the keyboard. Those who answered all "Y" or all "N" in any test block were excluded for failure to comply with the instructions.

After training and testing on one artificial language, participants in the Naïve condition were debriefed and given permission to leave, whereas participants in the Same Language condition and the 100/67%, 83/67%, and 67/67% Cued conditions began the second train-test phase. The procedure for this phase was the same as in the initial phase.

RESULTS

Preliminary analyses indicated that performance did not differ as a function of the language version (Version A vs. Version B) to which participants were exposed, and thus we collapsed across this factor in all subsequent analyses.

Phase 1 Performance

We first report the findings from Phase 1 for those groups given prior experience; the 100/67%, 83/67%, and 67/67% Cued groups, as well as the 67%/Same Language group. Following previous work using these materials (Lany et al., 2007), learning was assessed by creating a set of difference scores reflecting discrimination between grammatical strings and their respective ungrammatical foils to help account for the tendency to respond with "yes" to all strings. For each of the four test-string types, we subtracted the percentage of ungrammatical strings a participant endorsed (or false alarms) from their endorsement rates to the paired grammatical strings (or hits). Values above zero indicate that participants said "yes" more often to grammatical strings than to ungrammatical ones. Table 4 contains the mean difference scores for Phase 1 broken down by test string type and familiarization condition as well as the one-sample t tests (these and all subsequent comparisons were two-tailed, with alpha set to .05, Bonferroni corrected alpha for family-wise error rate of .0125). Inspection of Table 4 shows that participants in all conditions showed significant discrimination for Familiar +Feature Strings. Participants in the 100/67% Cued condition also showed significant discrimination for Generalization +Feature strings as compared to chance, but participants in the 83/67% and 67/67% Cued conditions and the 67%/Same Language condition did not. One sample t tests on the –Feature strings revealed that participants in the 83/67%-Cued, 67/67%-Cued, and 67%/Same Language conditions showed significant discrimination for Familiar strings but not for Generalization strings.

We next tested whether there were group differences in discrimination. Because participants in the 100/67%-Cued condition always heard strings containing features in Phase 1, we tested for group differences in performance for +Feature strings separately from testing for differences in performance on –Feature strings.

Group analyses on +Feature strings. Beginning with the +Feature strings, a mixed ANOVA, with familiarization condition as a between participant factor and test string type

TABLE 4
Phase 1 Test Performance

	Familiar +Feature		*Familiar –Feature*		*Generalization +Feature*		*Generalization –Feature*	
Familiarization Condition	*H-FA*	t *test*	*H-FA*	t *test*	*H-FA*	t *test*	*H-FA*	t *test*
100% Cued	.30 (.039)	$t(41) = 6.3^{***}$	n/a	n/a	.18 (.041)	$t(41) = 3.6^{***}$	n/a	n/a
83% Cued	.30 (.040)	$t(39) = 7.1^{***}$	.27 (.037)	$t(39) = 5.8^{***}$	.03 (.043)	$t(39) = .7$	.10 (.039)	$t(39) = 1.9$
67% Cued	.27 (.039)	$t(41) = 8.9^{***}$	.20 (.036)	$t(41) = 7.4^{***}$	.01 (.041)	$t(41) = .3$	.05 (.038)	$t(41) = 1.4$
67% Cued Same Language	.21 (.038)	$t(43) = 6.1^{***}$	.19 (.035)	$t(43) = 6.0^{***}$	.01 (.041)	$t(43) = .3$	.06 (.038)	$t(43) = 1.9$

*** = $p < .0125$; Note that $P = .05/4 = .0125$ meets the criterion for a Bonferroni correction for conducting 4 tests in each family (e.g., Familiar +Feature, Familiar –Feature, etc) to control for family-wise error rate. Performance scores reflect a Hits minus False Alarms (H-FA) difference score. Standard errors appear in parenthesis next to the corresponding means. One sample, two-tailed *t* test results using the H-FA measure as a difference score are shown in the table. Significant discrimination is marked with asterisks in the corresponding column. Note that participants in the 100% Cued condition were not exposed to strings without features.

(Familiar +Feature and Generalization +Feature) as a within participant factor, revealed better discrimination for Familiar +Feature test strings than for Generalization +Feature strings (M = .26, SE = .02, and M = .06, SE = .02, respectively), $F(1,164) = 109.96$, $p < .001$, $\eta_p{}^2 = .4$ reflecting a robust advantage of familiar over generalization strings. Critically, however, there was also an interaction between test-string type and familiarization condition, $F(3,164) = 2.83$, $p = .04$, $\eta_p{}^2 = .05$, reflecting different patterns of responding to Familiar and Generalization strings across familiarization conditions. A one-way ANOVA on Familiar +Feature test strings indicated that performance did not differ across the four familiarization conditions, $F(3, 164) = 1.3$, $p = .28$ (see Table 4). In contrast, there was a significant effect of familiarization condition for Generalization +Feature test strings, $F(3, 164) = 3.64$, $p = .014$, $\eta_p{}^2 = .06$. In a series of planned orthogonal comparisons, we tested the hypothesis that sensitivity to the marker-feature relationships decreases as a function of cue reliability. In line with this prediction, the 100% Cued condition performed better than the other three groups combined, $t\ (164) = 3.23$, $p = .001$, $d = .5$ on the Generalization +Feature strings. When we compared the 83% Cued condition to the two groups exposed to a 67% cued language (the 67/67% Cued and 67%/Same Language groups) they did not differ, $t\ (164) = .46$, $p = .64$, $d = .07$.

The results for the Familiar +Feature test strings suggest that participants in all conditions were equally able to recognize familiar strings containing the marker-feature co-occurrence relationships. However, for Generalization +Feature strings, which are a stronger test of learning of the marker-feature relationships because they have not been heard, participants benefitted from higher cue-probability. Learners exposed to strings in which *a*s always predicted words with a particular ending generalized to unfamiliar strings containing that regularity, while participants for whom some strings did not conform to this pattern did not (Table 4, column 3). Nonetheless, even participants in the 100/67% Cued condition did not endorse novel strings containing the marker-feature relationship to the same extent that they endorsed familiar strings, indicating that while they were sensitive to the co-occurrence relationships between markers and features, their sensitivity was greatest for strings that they had previously heard.

Group analyses on –Feature strings. We next examined the performance of the 83/67%-Cued, 67/67%-Cued, and 67%/Same Language conditions on –Feature strings using an ANOVA with familiarization condition as a between-participant factor, and test string type (Familiar –Feature and Generalization –Feature) as a within participant factor. The results revealed better discrimination for Familiar strings (M = .22, SE = .021) than for Generalization strings (M = .067, SE = .022), $F(1, 83) = 3.08$, $p < .001$, $\eta_p{}^2 = .27$. There were no other significant main effects or interactions.

In summary, in the two sets of analyses (+Feature and –Feature), there were no reliable differences among the 83/67% Cued, 67/67% Cued, and 67%/Same Language conditions. Moreover, while participants in all conditions discriminated familiar strings from ungrammatical ones, only the participants exposed to the 100% Cued language learned the marker-feature relationships as reflected in their performance on Generalization +Feature items. Participants exposed to an 83% Cued or a 67% Cued language also failed to generalize to new –Feature items (strings lacking the marker-feature relationships), in line with previous studies suggesting that learning the higher-level *aX bY* co-occurrence restrictions that are not dependent on a feature being present is quite difficult (Braine, 1987; Frigo & McDonald, 1998; Gerken et al., 2005).

Phase 2 Performance

Discrimination based on type of prior familiarization. To examine Phase 2 performance, we first tested whether participants in each familiarization condition showed significant discrimination for each test string type. Table 5 contains the means and standard errors broken down by Familiarization Condition and test string type, as well as the outcomes of one-sample *t* tests measuring significant discrimination (as compared to chance) for each kind of test trial. Consistent with the Phase 1 performance of the 67/67% Cued and 67%/Same Language participants, the 67% Naïve group showed significant discrimination for Familiar strings (+Feature and – Feature), while failing to show discrimination for Generalization strings (both +Feature and –Feature). Interestingly, participants in the 67%/Same Language and 100/67% Cued conditions showed the same pattern of performance as the Naïve participants: discrimination for Familiar strings (both + and –Feature), but no evidence of discriminating Generalization strings from ungrammatical ones. In contrast, participants in both the 83/67% and 67/67% Cued conditions showed significant discrimination for all test string types. These participants' successful discrimination for Generalization +Feature strings suggests they had learned the co-occurrence relationships between the *a*s and *b*s and the distinctive endings on the *X*s and *Y*s. However, the fact that they *also* showed discrimination for the Generalization –Feature strings suggests that beyond having learned the relationships between markers and features, they were sensitive to more abstract category-level co-occurrence relationships. Thus, in Phase 2, only participants in the 67/67% and 83/67% Cued conditions showed evidence of sensitivity to the abstract *aX bY* relationships; that *a*s predict one set of words, and that *b*s predict a different set.

Group differences in performance based on type of prior familiarization. We next examined group differences in participants' ability to learn a new 67% Cued language using a mixed ANOVA with familiarization condition (67% Naïve, 100/67% Cued, 83/67% Cued, 67/67% Cued, and 67%/Same Language) as a between participant factor, and test string familiarity (Familiar vs. Generalization) and test string type (+Feature vs. –Feature) as within participant factors. The analysis revealed better discrimination for Familiar strings ($M = .26$, $SE = .02$) than for Generalization strings ($M = .05$, $SE = .01$), $F(1, 205) = 227.83$, $p < .001$, $\eta_p^2 = .5$, and better discrimination for +Feature strings ($M = .18$, $SE = .02$) than for –Feature strings ($M = .14$, $SE = .02$), $F(1, 205) = 5.04$, $p = .03$, $\eta_p^2 = .02$. More importantly, there was an effect of Familiarization condition, $F(4, 205) = 3.73$, $p = .006$, $\eta_p^2 = .07$. Inspection of mean discrimination across the five familiarization conditions (collapsed across the different types of test trials) shows that performance was best in the 67/67% Cued condition ($M = .24$, $SE = .04$), followed in order by the 83/67% Cued Condition ($M = .19$, $SE = .04$), the 67%/Same Language ($M = .14$, $SE = .02$) Condition, the 100/67% Cued condition ($M = .13$, $SE = .03$), and finally, the 67% Naïve control condition ($M = .08$, $SE = .02$). In a series of orthogonal planned contrasts (2-tailed) we further investigated the source of this group difference.

One vs. two version exposure. Our first and broadest question was whether participants exposed to two different sets of co-occurrence relationships (the 100/67%, 83/67% and 67/67% groups) differed from participants who were exposed to just one set (the 67% Naïve and 67%/Same Language groups). This comparison revealed significantly better performance

TABLE 5
Phase 2 Test Performance on the 67% Cued Language

	Familiar +Feature		*Familiar –Feature*		*Generalization +Feature*		*Generalization –Feature*	
Familiarization Condition	*H-FA*	t *test*	*H-FA*	t *test*	*H-FA*	t *test*	*H-FA*	t *test*
100/67% Cued	.26 (.043)	$t(41) = 5.8^{***}$	.20 (.045)	$t(41) = 4.3^{***}$	.03 (.045)	$t(41) = .6$	.04 (.033)	$t(41) = 1.3$
83/67% Cued	.28 (.044)	$t(39) = 5.7^{***}$	.25 (.046)	$t(39) = 5.1^{***}$	.13 (.046)	$t(39) = 2.4^{*}$	.10 (.034)	$t(39) = 2.3^{*}$
67/67% Cued	.38 (.043)	$t(41) = 7.6^{***}$	.36 (.045)	$t(41) = 6.7^{***}$	.12 (.045)	$t(41) = 2.2^{*}$	.09 (.033)	$t(41) = 2.8^{**}$
67/67% Cued Same Language	.29 (.042)	$t(43) = 7.7^{***}$	.22 (.044)	$t(43) = 5.5^{***}$	.01 (.044)	$t(43) = .3$	.04 (.033)	$t(43) = 1.5$
67% Cued Naïve	.21 (.043)	$t(41) = 6.7^{***}$	.13 (.045)	$t(41) = 3.5^{**}$	.04 (.045)	$t(41) = 1.3$	−0.06 (.033)	$t(41) = -1.9$

* = $p<.05$, ** = $P<.01$, *** = $p < .001$; Note that $P<.01$ meets the criterion for a Bonferroni correction for conducting 5 tests in each family (e.g., Familiar +Feature, Familiar –Feature, etc.). Performance scores reflect a Hits minus False Alarms (H-FA) difference score. Standard errors appear in parenthesis next to the corresponding means. One sample *t* tests on H-FA in which participants showed significant discrimination are marked with asterisks in the corresponding column.

for participants given experience with two different languages ($M = .18$, $SE = .02$) than one language ($M = .11$, $SE = .01$), t (205) = 2.7, $p = .008$, $d = .38$.We next directly compared the 67% Naïve with the 67%/Same Language participants to determine whether the 67%/Same Language participants benefitted from additional experience with the language ($M = .14$, $SE = .02$) relative to the 67% Naïve controls ($M = .08$, $SE = .02$), and found no difference, t (205) = 1.3, $p = .168$, $d = .18$.

Deterministic vs. probabilistic language exposure. We next asked whether experience with deterministic co-occurrence relationships (the 100/67% Cued condition) affected subsequent learning of a 67% Cued language differently than experience with probabilistically cued contingencies (the 83/67% and 67/67% Cued conditions). The 83/67% and 67/67% Cued groups significantly outperformed ($M = .24$, $SE = .04$) the 100% Cued group ($M = .14$, $SE = .02$) in Phase 2, t (205) = 2.1, $p = .037$, $d = .3$. There was no difference in performance between the 83/67% and 67/67% Cued groups, t (205) = 1.16, $p = .24$, $d = .16$, ($M = .19$, $SE = .04$ vs. $M = .24$, $SE = .04$, respectively).

Specific benefits resulting from prior exposure to probabilistically-cued co-occurrence relationships. Altogether these findings suggest that prior experience with a probabilistically-cued co-occurrence relationships provided the greatest overall benefit to learning novel probabilistically-cued relationships. We next directly compared the 83/67% and 67/67% Cued groups with the 67% Naïve learners using two ANOVAs with familiarization condition as a between participants factor and test trial type as a within participants factor. One comparing 67% Naive learners with the 67/67% Cued group, and the other comparing the 67% Naïve learners with the 83/67% Cued group. These analyses were necessary to determine whether participants in each group had an advantage over 67% Naïve participants.

A significant main effect of group indicated that the 67/67% Cued group performed better ($M = .24$, $SE = .04$) than the 67% Naïve controls ($M = .08$, $SE = .02$), $F(1, 82) = 13.46$, $p < .001$, $\eta_p{}^2 = .14$. Additionally, we found no significant interactions between familiarization condition and test trial type, suggesting that the 67/67% Cued participants performed better than 67% Naïve Controls for all test trial types (see Table 5 for mean performance broken down by trial type). Planned t tests comparing the two groups' performance on each type of test trial generally confirmed this picture, revealing an advantage for the 67/67% Cued condition for Familiar +Feature strings, t (82) = 2.84, $p = .006$, Familiar -Feature strings, t (82) = 3.63, $p < .001$, and, critically, for Generalization -Feature strings, t (82) = 3.33, $p = .001$. The 67/67% Cued participants' numerical advantage for Generalization +Feature strings failed to reach significance, t (82) = 1.28, $p = .2$).

When comparing the 83/67% Cued group and 67% Naïve controls, we found a significant main effect of group, with greater overall performance for the 83/67% Cued group ($M = .19$, $SE = .04$) than Naïve controls ($M = .08$, $SE = .02$), F (1,80) = 6.14, $p = .015$, $\eta_p{}^2 = .07$. There were no significant interactions between familiarization condition and test trial type. Planned t-tests comparing the two conditions on each kind of test trial revealed greater performance for the 83/67% Cued condition on Familiar -Feature strings and Generalization -Feature Strings, ts (80) $\geq$2.0 and ps $\leq$ .05 (see Table 5 for means and standard errors).

In sum, the 67/67% and 83/67% Cued groups each performed significantly better than the 67% Naïve group overall. The 67/67% Cued participants also showed more consistent advantages when performance was examined separately by trial type. However, neither the 67/67% nor the 83/67% groups showed significantly better performance than 67% Naïve controls on all trial types.

Within-participant changes in performance. Examining within-participant change provides an additional opportunity to assess the effects of experience on learning, and thus we tested changes in learning from Phase 1 to Phase 2 in participants exposed to two different language versions. We found that participants in the 100/67% Cued did not differ in their discrimination for Familiar items from Phase 1 to 2 (Phase 1 $M = .30$, $SE = .047$, and Phase 2 $M = .26$, $SE = .046$: $t\,(41) = .78$, $p = .4$) but had an advantage for Generalization items in Phase 1 vs. Phase 2 (Phase 1 $M = .18$ $SE = .049$, and Phase 2 $M = .03$, $SE = .052$: $t\,(41) = 2.87$, $p = .007$: the interaction between phase and test-trial type was significant, F $(1, 41) = 4.47$, $p = .04$. Because the marker-feature contingencies were substantially less reliable in Phase 2 relative to Phase 1 for this group, it is unclear whether it is reasonable to expect equivalent learning of these contingencies in a more probabilistic language. However, the fact that their performance on these contingencies in Phase 2 was not above chance ($M = .03$, $SE = .052$), suggests that they did not show strong learning of these contingencies in Phase 2.

For participants in the 83/67% Cued conditions, we found no change in learning between Phases 1 and 2 $F\,(1, 43) = .73$, $p = .4$. However, for participants in the 67/67%-Cued condition, performance was better in Phase 2 ($M = .24$, $SE = .039$) than Phase 1 ($M = .13$, $SE = .014$); F $(1, 41) = 6.75$, $p = .013$). There were no reliable differences in 83/67% and 67/67% Cued participants' level of performance in Phase 1, and thus the fact that only the 67/67% Cued participants showed an improvement over Phase 1 suggests that prior experience in this condition may provide the strongest foundation for subsequently learning co-occurrence relationships with matched cue levels.

GENERAL DISCUSSION

The current experiment investigated the effects of prior experience on learning probabilistically cued patterns. Participants were exposed to an artificial language containing co-occurrence relationships between word categories similar to grammatical dependencies such as the predictive relationships in English between determiners and nouns, and auxiliaries and verbs. In accord with previous studies, our findings suggest that these contingencies can be very difficult to learn if they are not reliably marked by correlated cues: In the absence of any prior experience, participants who were familiarized to a 67% or 83% Cued language successfully recognized the strings they had heard during familiarization but failed to learn anything about the co-occurrence relationships between categories (i.e., they learned neither the relatively concrete marker-feature relationships nor the more abstract *aX bY* category co-occurrence relationships). As in previous studies, we found that participants successfully learned the marker-feature co-occurrence relationships when 100% of *X*s and *Y*s contained the distinctive features, suggesting that highly reliable phonological cues marking *X*s and *Y*s can facilitate learning. However deterministic cues are a rarity in language.

In spite of the difficulty initially posed by the 83% and 67% Cued patterns, once exposed to them, participants showed superior overall learning of novel 67%-Cued co-occurrence relationships. These two groups' test performance was significantly better than Naïve learners' who lacked any prior experience. In contrast, although participants exposed to the 100% Cued language were the only ones to learn the marker-feature co-occurrence relationships in the initial training phase, this learning did not facilitate subsequent acquisition of novel, probabilistically cued co-occurrence relationships. Also, participants in the 67%/Same Language condition did not benefit from their prior experience with the *same exemplars* from the probabilistic pattern. While additional exposure often leads to better learning, these findings suggest that additional experience with a small set of items may not result in advantages to learning abstract structure. This pattern of findings also rules out the possibility that the structural properties of language stimuli are inherently represented in terms of abstract rules irrespective of input properties (Hypothesis 4 in the introduction). If this were the case, there should be no differences between the 67%/Same Language and 67/67% Cued conditions. Furthermore, these findings suggest that reducing processing demands through additional exposure cannot explain the superior performance of learners in the 83/67% and 67/67%-Cued conditions (Hypothesis 3).

Altogether, these data suggest that experience with dissimilar surface features can play a central role in learning probabilistically cued co-occurrence relationships (Hypothesis 2). Natural languages incorporate patterns that differ both in their surface features and in the cue-reliability of these features but contain similar underlying structure (e.g., determiner-noun and auxiliary-verb co-occurrence relationships), and these findings suggest that learning such abstract language structure may be supported by gaining experience with variable surface instantiations of a pattern. However because participants in these conditions showed no evidence of learning the marker feature co-occurrence relationships or the more abstract category co-occurrence relationships in Phase 1, it is important to consider how their prior experience promoted subsequent learning. The fact that they successfully discriminated familiar strings from ungrammatical ones in Phase 1 indicates that they were encoding information about the strings they heard. We suggest that experience with a new language (in terms of vocabulary and/or probabilistically cued contingencies) led learners to notice some of the similarities between the strings in the two languages. For example, they may have noticed that in both languages, strings frequently began with one of four short words and ended in one of two syllables. Because tracking co-occurrence relationships between markers and features is thought to be a critical component of category learning, enhanced attention to those aspects of the language may have begun to clue participants in to the *aX bY* structure. This explanation would also hold if participants had begun to learn the marker-feature co-occurrence relationships in Phase 1, but not well enough to reliably discriminate the Generalization strings from ungrammatical ones. Experience with a new language with similar features would likewise encourage participants to track the underlying structure shared by both languages more closely, leading to successful learning. This account is consistent with the theory that benefits from prior experience arise as learners' attention is trained to relevant dimensions of stimuli (e.g., Smith et al., 2002).

However, not all forms of prior experience with different surface structure appear to promote learning. Prior experience with a probabilistically cued pattern (i.e., the 83% or 67% Cued) promoted sensitivity to novel 67% Cued co-occurrence relationships, but surprisingly, prior experience with a perfectly cued, or deterministic, pattern failed to facilitate subsequent learning, even though it resulted in the best learning initially. Thus, while the 100% and 67% cued languages in Phases 1 and 2 both involved adjacent co-occurrence relationships between word categories, there appear to be important differences in how learners responded to them. An intriguing possibility is suggested by findings that experience with language-wide patterns can dramatically influence processing of novel sentences. Wonnacott, Newport, and Tannenhaus (2008) exposed adults to an artificial language in which verbs could occur in two different sentence constructions. When most verbs occurred in both constructions, learners exposed to a novel verb in just one construction showed evidence of expecting that it could occur in the other construction, despite the absence of any explicit positive evidence. In contrast, when exposed to a language in which most verbs occurred in only one of two possible constructions, learners exposed to a novel verb in just one construction rated instances of that verb in the alternate construction more poorly. These results suggest that learners respond differently to the statistics of specific items as a function of what they already know about the language-wide statistical properties of their language. The current study also suggests that language-wide statistics can impact how the same statistical regularities influence learning novel structures; in this case, experience with a deterministic pattern may have changed how participants responded to a probabilistic (but still reliable) pattern within the same experiment. Exposure to a completely deterministic pattern may have prevented participants from noticing the probabilistic determiner-feature relationships in the second phase, or skewed their weighting of those contingencies. It also may have led them to focus only on the specific strings they experienced in Phase 2, on which they excelled. In contrast, as described above, participants in the 83/67% and 67/67 % conditions may have noticed that *a*s were followed by words with particular phonological features more often than not, and this sensitivity could have tuned them in to similar predictive features in the novel co-occurrence relationships.

Another potential explanation for this finding is that learning probabilistic and deterministic patterns are largely subserved by different underlying mechanisms, as has been claimed in studies investigating the output of learning whether it taps statistics versus rules. For example, Pena, Bonatti, Nespor, and Mehler (2002) found that adults used reliable transitional probabilities between nonadjacent syllables to segment words in a continuous speech stream. Learners did not, however, generalize, to novel words that maintained the nonadjacent dependencies but contained a novel middle syllable. However, when the syllable stream was segmented by brief pauses, presumably eliminating the need to track transitional probabilities for segmentation purposes, adults both learned the nonadjacent dependencies and generalized to novel instances containing that structure (see also Endress & Bonatti, 2007), exhibiting something akin to rule learning. The authors interpret these findings as evidence that different mechanisms are involved in segmenting words via statistical information and forming abstract rules about word-internal structure. Although there are alternate accounts of these data that account for these effects within a single learning system (e.g., Perruchet, Tyler, Galland, & Peereman, 2004), recent studies investigating the neural mechanisms involved in word segmentation versus abstracting structural properties of words suggest that these processes may differ (Cunillera et al., 2009; Cunillera, Toro, Sebastian-Galles, & Rodriguez-Fornells, 2006; De Diego Balaguer, Toro, Rodriguez-Fornells, & Bachoud-Levi, 2007). Learning sequences generated by an artificial grammar may also rely on different mechanisms depending on whether the sequences recruit primarily explicit or implicit learning mechanisms (Destrebecqz et al., 2005). Whatever the neural basis underlying learning of the deterministic and probabilistic patterns in the current study, they too may rely on different neural systems. If this is the case, then changes in the system used for deterministic learning might not extend to the system involved in learning the probabilistic language. Interestingly, similar findings have been reported in other domains of learning, not just language. Neuropsychological and neuroimaging studies of category learning outside the realm of language suggest that the neural processes supporting learning depend on the nature of the category (Ashby & Spiering, 2004). When categories are probabilistically cued by a set of features, as in the weather-prediction task developed and extensively studied by Gluck and colleagues (e.g., Gluck & Bower, 1988), learning seems to rely more heavily on the basal ganglia and striatum (see Shohamy, Myers, Kalanithi, & Gluck, 2008, for a review) than when the categories can be distinguished by a relatively simple dimension, as in the Wisconsin Card Sorting Task.

While the current findings are intriguing, we should note several limitations on their interpretation. In the current experiment, we found that prior experience can enhance subsequent learning in a novel domain, but it will be important in future studies to test whether the specific findings demonstrated here hold for infants. Like adults, infants successfully learn an *aX bY* language with the support of strong correlated cues (Gerken et al., 2005; Gómez & Lakusta, 2004). Infants often fail to generalize their learning to new instances that are low in perceptual similarity, but there are noteworthy exceptions to this trend. Specifically, infants have shown evidence of generalizing sensitivity to novel vocabulary in other artificial-language learning tasks (Marcus et al., 1999; Gómez & Gerken, 1999). Thus, it is an open question whether sensitivity to an *aX bY* language will affect processing of novel exemplars following that pattern in infants.

Additionally, while our artificial language was quite challenging for participants, it is simple in comparison to natural languages. Thus, it will be important to begin to test the predictions arising from these studies under conditions that can scale up to those encountered when learning natural language. One possibility would be to design studies that test predictions arising from our findings in the natural course of language acquisition. For example, we might test whether learning determiner noun co-occurrence relationships reliably precedes the emergence of sensitivity to other co-occurrence relationships (e.g., pronoun verb co-occurrence relationships), and whether mastery of multiple such relationships coincides with the emergence of sensitivity to more abstract levels of this structure. A related issue is that while learners may use an early-acquired sensitivity to bootstrap sensitivity to a related pattern in natural language, it is highly unlikely that the occurrence of the structures would also be sequential and nonoverlapping in the input. We exposed learners to sequential, nonoverlapping input to cleanly assess the effects of one structure on the other, but future studies should examine how simultaneous exposure affects this learning.

In sum, the current study suggests that experience with patterns that vary in their surface features but have similar underlying structure plays an important role in developing an abstract sensitivity to the category-level co-occurrence relationships. These findings shed new light on the mechanisms by which we learn probabilistically cued co-occurrence relationships between word categories, a critical task in natural language learning. The findings also underscore the important role that experience plays in shaping learning over the course of language acquisition, selectively tuning learners to relevant structure in their language input.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Grant F32 HD057698 to J.L. and R01 HD42170 to R.L.G. We thank Jessica Payne for thoughtful comments on an earlier version of this paper.

REFERENCES

Ashby, F. G., & Spiering, B. J. (2004). The neurobiology of category learning. *Behavioral and Cognitive Neuroscience Reviews*, *3*, 101–113.

Braine, M. D. S. (1987). What is learned in acquiring words classes: A step toward acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 65–87). Hillsdale, NJ: Erlbaum.

Christiansen, M. H., Onnis, L., & Hockema, S. A. (2009). The secret is in the sound: From unsegmented speech to lexical categories. *Developmental Science*, *12*, 388–395.

Coluga, E., & Smith, L. B. (2003). The emergence of abstract ideas: Evidence from networks and babies. *Philosophical Transactions of the Royal Society*, *358*, 1205–1214.

Coluga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, *112*, 347–382.

Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., . . . Rodríguez-Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *NeuroImage*, *48*, 541–553.

Cunillera, T., Toro, J. M., Sebastian-Galles, N., & Rodriguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Research*, *1123*, 168–178.

De Diego Balaguer, R., Toro, J. M., Rodriguez-Fornells, A., & Bachoud-Levi, A. (2007). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS ONE*, *11*, e1175.

Destrebecqz, A., Peigneux, P., Laureys, S., Degueldre, C., Del Fiore, G., Aerts, J., . . . Maquet, P. (2005). The neural correlates of implicit and explicit sequence learning: Interacting networks revealed by the process dissociation procedure. *Learning and Memory*, *12*, 480–490.

Devescovi, A., Caselli, M. C., Marchione, D., Pasqualetti, P., Reilly, J., & Bates, E. (2005). A crosslinguistic study of the relationships between grammar and lexical development. *Journal of Child Language*, *32*, 759–786.

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous stream. *Cognition*, *105*, 247–299.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 234–257.

Frigo, L., & McDonald, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language*, *39*, 218–245.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, *52*, 45–56.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition, 65*, 263–297.

Gentner, D., & Namy, L. (1999). Comparison in the development of categories. *Cognitive Development*, *14*, 487–513.

Gerken, L. A., Wilson, R., & Lewis, W. (2005). Seventeen-month-olds can use distributional cues to form syntactic categories. *Journal of Child Language*, *32*, 249–268.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology—General*, *117*, 227–247.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431–436.

Gómez, R. L., & Gerken, L. A. (1999). Artificial grammar learning by one-year-olds leads to specific and abstract knowledge. *Cognition*, *70*, 109–135.

Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, *7*, 183–206.

Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction in 12-month-old infants. *Developmental Science*, *7*, 567–580.

Jones, S. S., & Smith, L. B. (2002). How children know the relevant properties for generalizing object names. *Developmental Science*, *5*, 219–232.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, *99*, 349–364.

Lany, J. (2012). *Infants' use of probabilistic phonological and distributional regularities marking grammatical categories in word learning: Relationships to language proficiency*. Manuscript in preparation.

Lany, J., & Gómez, R.L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science*, *19*, 1247–1252.

Lany, J., Gómez, R. L., & Gerken, L. A. (2007). The role of prior experience in language acquisition. *Cognitive Science*, *31,* 481–507.

Lany, J., & Saffran, J. R. (2010). From statistics to meaning: Infants' acquisition of lexical categories. *Psychological Science*, *21*, 284–291.

Lany, J., & Saffran, J. R. (2011). Interactions between statistical and semantic information in infant language development. *Developmental Science*, *14*, 1207–1219.
Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*, 77–80.
Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, *26*, 393–424.
Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, *55*, 259–305.
Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorization. *Cognition*, *96*, 143–182.
Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*, 185–200.
Pena, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*, 604–607.
Perruchet, P., Tyler, M. D., Galland, N., & Peermen, R. (2004). Learning nonadjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology, 133,* 573–583.
Pizzuto, E., & Caselli, M. C. (1992). The acquisition of Italian morphology: Implications for models of language development. *Journal of Child Language*, *19*, 491–557.
Shohamy, D., Myers, C. E., Kalanithi, J., & Gluck, M. A. (2008). Basal ganglia and dopamine contributions to probabilistic category learning. *Neuroscience and Biobehavioral Reviews*, *32*, 219–236.
Smith, K. H. (1969). Learning co-occurrence restrictions: Rule learning or rote learning? *Journal of Verbal Behavior*, *8*, 319–321.
Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, *13*, 13–19.
Tomasello, T., & Olguin, R. (1993). Twenty-three-month-old children have a grammatical category of noun. *Cognitive Development, 8,* 451–464.
Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature artificial language. *Cognitive Psychology*, *56*, 165–209.