There is a substantial literature describing how infants become more sensitive to differences between native phonemes (sounds that are both present and meaningful in the input) and less sensitive to differences between non-native phonemes (sounds that are neither present nor meaningful in the input) across development. Here, we review an emergent strand of literature that gives a more nuanced notion of the problem of sound category learning. This research documents infants’ discovery of phonological status, signaled by a decrease in sensitivity to sounds that map onto the same phonemic category vs. different phonemic categories. The former phones are present in the input, but their difference does not cue meaning distinctions because they are tied to one and the same phoneme. For example, the diphthong /i/ in “I’m” should map to the same underlying category as the diphthong in “I’d,” despite the fact that the first vowel is nasal and the second oral. Because such pairs of sounds are processed differently than those that map onto different phonemes by adult speakers, the learner has to come to treat them differently as well. Interestingly, there is some evidence that infants’ sensitivity to dimensions that are allophonic in the ambient language declines as early as 11 months. They lay out behavioral research, corpora analyses, and computational work which sheds light on how infants achieve this feat at such a young age. Collectively, this work suggests that the computation of complementary distribution and the calculation of phonetic similarity operate in concert to guide infants toward a functional interpretation of sounds that are present in the input, yet not lexically contrastive. In addition to reviewing this literature, we discuss broader implications for other fundamental theoretical and empirical questions.

Keywords: infants, perception, phonology, phonemes, speech
literature that documents when infants begin to apply a differential processing of allophones and phonemes (Section "Infants' processing of allophones") and how they might have learned to make such a distinction between allophones and phonemes (Section "Mechanisms for learning allophones"). The final section ("Implications") discusses how research on infants' learning of phonological status can inform, and be informed by, other areas of investigation. Throughout this article, we identify areas where answers are still lacking. We hope that this review serves as a springboard for such work and helps to point to clear areas out of which this future work can grow.

WHAT ARE ALLOPHONES?

There are two classical cases of allophony, which are commonly discussed in introductory linguistics courses (Trubetzkoy, 1939/1969; Kenstowicz, 1994). The first involves "sounds in complementary distribution." Two sounds are in complementary distribution if the sound which should be used is completely predictable from the context; put differently, the contexts in which each sound can occur are completely non-overlapping. For example, in most varieties of American English, dark /l/ occurs syllable-finally ("ball"), whereas light /l/ occurs in all other positions ("lab"). Notice that no two words in American English differ only on whether they have a light or dark /l/. In other words, sounds in complementary distribution do not cue meaning distinctions. Finally, a third criterion for allophony in this case is that the two sounds must be somehow acoustically related, such that they may be interpreted as the "same" sound, on some abstract level. For instance, although /h/ and /f/ are in complementary distribution in English (the former occurs only in syllable codas, the latter only in syllable onsets), phonologists would not want to posit that they are allophones since they are highly acoustically distinct (Bazell, 1954).

The second classical case of allophony relates to sounds in "free variation." In this case, speakers can produce two or more different sounds in the exact same environment (e.g., ri[
]er versus ri[d]er in American English); however, these differences are not lexically relevant. Much work debates the name "free," since in many such cases the variant which is selected appears to be explained, to a considerable extent, by a number of structural, sociolinguistic, and idiolectal variables (e.g., Fischer, 1958). Nonetheless, it remains the case that two sounds which can be thus exchanged without semantic changes can be viewed as allophones. The traditional way of establishing whether two sounds are in free variation is by carrying out a minimal pair test. Minimal pairs are two wordforms that differ only in one sound; if this sound swap results in meaning change or loss, then the two sounds are phonemes, but if it does not, they are allophones in free variation.

In phonology, as in life, things can sometimes get more complicated, and for the ambiguous case from Scottish Standard English, where some vowels have long and short variants that are contextually determined, yet for which some minimal pairs, with specific morphological characteristics, can nevertheless be found. This is the case for long and short variants of /ai/, which contrast minimally in "side" and "sighed," with the long version being found in morphologically complex items. In spite of the existence of such minimal pairs, the two sounds are in free variation across speakers in some lexical items, such as "crisis."

In view of such cases both within and across languages, Pierre-humbert (2003) proposes to do away with the distinction between phonemes and allophones and instead attempt an explanation of learners' acquisition of positional allophones, defined as clusters of tokens in acoustic space. A comparable proposal was made in Ladd (2006), who goes further by pointing out that allophones are sometimes very meaningful sociolinguistically, and are thus highly perceptually salient to native speakers. Scobbie and Stewart-Smith (2008) argue, instead, that while the concepts of allophones

---

2 One anonymous reviewer points out that this problem only exists in the case that the speech stream is segmented, since only segmentation into syllables would lead to the conclusion that the allophone of /lu/ that occurs in "st" clusters does not group with the phonetically similar /d/; but rather with the phonetically dissimilar /t/. It is unclear when exactly infants segment into syllables. Some data point to syllables being the basic unit of analysis even for newborns, allowing the discrimination of /atpa/-apa/ but not that of /tspa/-pat/ (Bertoncini et al., 1988); while other data suggests a protracted development, as infants do not use the syllable-determined allophones of /te/ in "night rate" versus "nitrate" to segment these words from running speech until about 10 months of age. Thus this is certainly a question that warrants further exploration.

3 This sort of "active" contrast is eminently common in sign languages, which have very few clear minimal pairs. Thus while a few clear minimal pairs exist e.g., in the domain of handshape (Brentari, 1996), most are cases of near-minimal pairs or cases where a contrast exists in one area of the lexicon, but looks distinct in another area of the lexicon (Brentari and Eccarius, 2012). For example, according to Diane Brentari (p.c.) the ASL sign THOUSAND was originally borrowed from the initial French Sign Language sign MILLE and had a 3-finger "M" handshape. During the process of nativization the "M" (3 fingers) became "B" (all 4 fingers). That is the more marked 3 fingers became the less marked 4 fingers handshape. However, in this location with this movement there are no minimal pairs with either of these handshapes.
and phonemes may be useful end points, a continuum could exist between allophones and phonemes, and propose that speakers/listeners' grammars could well be fuzzy. More recently, Hall (2009) makes specific proposals as to how to predict perceptibility from gradient versions of an allophony/phonemicness scale. Clearly the limitations of the classical definitions of allophones and phonemes are not new (see e.g., Pike, 1947), but they are just now beginning to gain a unique combination of linguistic and psycholinguistic attention as it becomes increasingly clear that such phenomena are not marginal, and that such gradience is relevant to both language learning and processing. Indeed, a look through Table 1 reveals a window into the scope of this gradience. While it is not the aim of this paper to debate phonological theory, nor to enumerate cases along this continuum, we keep the question of gradience in mind when considering how infant learners may approach the phonological system, and what types of allophones versus phonemes (i.e., at what point of the continuum) have been studied in previous experimental work. With this enriched view of allophony, we now turn to adults' perception of these two (or more) "classes" of sounds.

THE END STATE: ADULTS' PROCESSING OF ALLOPHONES

It should be noted from the outset that the study of whether listeners' sensitivity for contrasts that are allophonic is lower than those that are phonemic faces certain methodological roadblocks, which are worthy of discussion here. One way to interpret this hypothesis is the following: Holding the listener and language constant, one would compare a contrast A that is phonemic against a contrast B that is allophonic, to find that A is processed better (discriminated more speedily and accurately; used for tracking phonological patterns; recruited for coding lexical contrasts). Much of the initial literature we review uses this design (e.g., Whalen et al., 1997; McLennan et al., 2003). When using this design, there is an obvious interpretation alternative to phonological status affecting perception: perhaps there is an intrinsic discriminability difference between A and B. A safeguard against this state of affairs is to test two sets of participants, who have different native languages, and hold the contrast constant, an approach that is also common in the literature (e.g., Johnson and Babel, 2010). In this scenario, intrinsic differences in discriminability of contrasts are irrelevant, since only one contrast is used. However, another problem arises, namely that the stimuli must be recorded in some language. If they are recorded in the language where the contrast is allophonic, they may be pronounced less clearly (provided that speakers tend to neutralize such contrasts), which is not desirable. But if they are recorded in the language where they are phonemic, then the test may facilitate the performance of listeners of that language, who will find the stimuli native. The solution that researchers are increasingly adopting is to use a third language where the contrast is phonemic, such that the stimuli are equally foreign to both sets of participants. Results from the latter approach actually fit in perfectly with conclusions derived from the two other (e.g., Boomershine et al., 2008), lending further credence to this body of literature.

In brief summary, previous work suggests that adults do not discriminate allophonic alternates as well as phonemes both behaviorally (Whalen et al., 1997; Peperkamp et al., 2003; Boomershine et al., 2008; Shea and Curtin, 2011) and electrophysiologically (Kazanina et al., 2006; Hacquard et al., 2007). Furthermore, adults rate allophones as more similar to each other than phonemes (Johnson and Babel, 2010). Additionally, words differing in sounds that are allophonic prime each other, but words differing in sounds that are phonemic do not (McLennan et al., 2003). These differences in processing come about as the result of native language exposure and thus second language learners have a hard time gaining sensitivity to sounds that are phonemic in the target language even when those same sounds are present allophonically in the learners' native language (Kondaurova and Francis, 2008).

Thus, overall, perceptual evidence in adults confirms that allophonic and phonemic sounds are not treated similarly. Given that there may be a continuum of allophones to phonemes, as mentioned above, it is worthwhile to evaluate whether this differential behavior arises only for the categorical phonemic/allophonic stages, or also for intermediate cases. This is especially true given recent findings that listeners treat sounds differently based on the reliability of their distributions (Dahan et al., 2008). Specifically, in this study Dahan et al. (2008) examined adults' perception of tensed and laxed variants of /æ/ in the environment of /k/ and /g/. When /æ/ is consistently tensed before /g/, but not /k/ they found a training effect in the experiment such that listeners, upon hearing e.g., the non-tensed /æ/ came to anticipate the following segment as /k/. Thus, when segments vary allophonically in a reliable way this can lead to differential processing very quickly. This is not the case when the variation is not predictable.

In Table 1, we reclassify adult perceptual studies in terms of the type of contrast that has been examined. There are several studies which explore one of the endpoints (e.g., Whalen et al., 1997), and other laboratories have examined a case of clear complementary distribution) and only a few studies which explore points in between. For example, in English [ɾ] can never map onto /d/, and therefore they form a phonemic contrast. Whereas [ɾ] is a possible realization of /d/ in word-medial context, there are also quite a few lexical exceptions where they occur in near overlapping distributions (e.g., rider vs. writer). Thus, the comparison of English adults' perception of the [ɾ] and [d], on the one hand, against [ɾ] and [d], on the other, represents the study of an intermediate case of allophony. This was undertaken in Boomershine et al. (2008), who found poorer discrimination of the former than the latter. Results cannot be attributed to the actual acoustic items used, since the same mapped onto different types for a second group of adults, whose native language was Spanish. In Spanish [ɾ] and [d] are in complementary distribution (classic allophony) and [ɾ] and [ɾ] are mostly in overlapping distribution (classic phonemic). Despite the fact that not all items fell on the extremes of the phonemicness/allophony continuum, perceptual results were the opposite across language groups in all tests but a measure of reaction time. Thus, this work seems to suggest that differences between phonemic and allophonic processing are found even when non-extreme points of the continuum are investigated.

Nonetheless, a different pattern emerges from work using electrophysiology. Hacquard et al. (2007) and Kazanina et al.
INFANTS’ PROCESSING OF ALLOPHONES

As mentioned above, a considerable body of literature suggests that perception of non-native (absent) sounds declines, whereas perception of native phonemes improves toward the end of the first year of life (Polka et al., 2001; Kuhl et al., 2006; Narayan, 2006). This improvement likely relates to the much richer and more abundant cues for the former: The child will accumulate more passive, phonetic exposure to the former; she may attempt these sounds; she may learn some words that have them, and so forth. The first question we would like to answer is when listeners become less sensitive to allophonic distinctions and more sensitive to phonemic ones. We review evidence from discrimination, phonotactic learning, phonotactic processing, and word learning suggesting that infants are sensitive to phonological status.

English-learning 2-month-olds discriminate allophonic variants (e.g., /t/ in “night rate” versus “nitrates”; Hohne and Jusczyk, 1994) showing an initial sensitivity to sounds that will eventually be treated as allophones later in life. Recent work suggests that, while young infants are sensitive to sounds that are allophones in their ambient language, this sensitivity declines with maturation and language-specialization. Specifically, Seidl et al. (2009) briefly familiarized English- and Quebec French-learning infants with a pattern that depended upon vowel nasality. Note that as mentioned earlier vowel nasality is phonemic in French, but allophonic in English. Infants in this study heard syllables in which nasal vowels were followed by fricatives, but oral vowels were followed by stops. Then they were tested on their ability to generalize this pattern to new syllables. English-learning 4-month-old infants were able to learn this novel phonotactic dependency involving vocalic allophones and behaved like French-learning 11-month-old infants, for whom nasality is phonemic. However, by 11 months of age English-learners were no longer able to encode this abstract phonotactic regularity and showed no evidence of learning. It should be noted that these older infants are not completely impervious to allophones, since they use them to extract words from running speech at 10.5 months (Jusczyk et al., 1999). Rather, these results suggest that the same exact sounds no longer function in the same manner across languages which use them as phonemes versus allophones.

It might be suggested that some of the contrasts that have been studied as allophones could be more perceptually difficult than ones that have been explored as phonemes. Specifically,
allophonic alternates may simply be more difficult to discriminate because they represent subtle changes. For example, Pegg and Werker (1997) found that two phones that map onto different phonemes /t/ and /d/, but are extremely similar, are not discriminable by one-year-olds. In their study, they measured sensitivity to the word-initial realization of /d/ against the post-/s/ realization of /t/, which differ very subtly. However, an important point is that simple acoustic distance between the tokens used in any given test cannot explain developmental changes in allophonic sensitivity, since this sensitivity changes with age and language exposure. Even in the Pegg and Werker (1997) study, 6-month-old were, in fact, able to distinguish the very similar surface realizations of /t/ and /d/. Similarly, Dietrich et al. (2007) and Seidl et al. (2009) show that attention to the same contrast declines in languages for which they are allophonic, but not in languages in which they are phonemic. For example, Dietrich et al. (2007) show that 18-month-old Dutch-, but not English-learning toddlers interpret vowel length as lexically contrastive. Thus, it appears that while sensitivity to allophonic sounds initially exists in infancy, it appears to decline by 11 months of age (Seidl et al., 2009) as infants converge on the native phonemic contrasts present in their input language and come to ignore the non-native ones which are not present in their input language (Werker and Ties, 1984).

It is worthy of note that most of the studies cited above have been conducted on English-learning infants (albeit with two exceptions, Dietrich et al., 2007; Seidl et al., 2009). If we are to draw any clear conclusions concerning the time course of allophonic sensitivity, we will need to expand this work cross-linguistically, since it may be that the time course is different across languages and may also be impacted by the kind of sound distinction explored. Unfortunately, such single language studies only allow for certain allophonic sounds to be tested, and confound potential differences in discriminability with phonological status.

Also worthy of mention is that many allophonic alternates in the studies mentioned above are predictable from the phonological context. For example, the aspiration of /t/ studied in Hohne and Jusczyk (1994) represents a clear case of complementary distribution or classic allophony. Exceptions to this are the cases of vowel nasalization utilized in Seidl et al. (2009) and the case of vowel length in Dietrich et al. (2007). Specifically, although vowels are nasalized before tautosyllabic nasal consonants in English, they are also often nasalized in other locations (e.g., within a word with another nasalized vowel), so complementary distribution does not entirely hold. Thus, although there are cases where nasalization of vowels is completely predictable on phonotactic grounds (before nasal Cs in the same syllable), we also see nasalization in other locations for coarticulatory reasons. To add more complexity to this picture, variation in nasalization has been reported across American English dialects, such that nasalization could become a sociolinguistically relevant feature (e.g., a marker of African American Vernacular English), more than a phonotactically relevant one. Similarly, in Dutch we see a case where vowel length is difficult to classify using our classic definition of allophony. Although there are minimal pairs with vowel length in Dutch, the presence if minimal pairs occurs unevenly across the inventory. For example, /j/ has long and short minimal pairs that differ mostly in length (although there are slight vowel quality differences). All other vowels that have been described as contrastive in length show considerable changes of vowel quality with the addition of length, much as we see in English tense-lax pairs. Certainly, the infant literature is not rich enough to conclude that there are no differences among the different degrees of allophony. Nonetheless, current research suggests that in infants, as in adults, even degrees of contrastiveness may make a difference, with more allophonic pairs being processed less well than more phonemic pairs. Across all studies, however, it appears that younger infants attend to salient distinctions more than older infants when the distinctions are allophonic in the target language.

MECHANISMS FOR LEARNING ALLOPHONES

Young toddlers treat allophones as distinct from phonemes. Further, some of the evidence reviewed suggests that they come to do so within the first year of life. How does such a young toddler come to treat allophones as distinct given that they clearly vary from language to language? Or more specifically, how do they come to attend less to allophonic sound pairs and attend more to phonemic sound pairs? There are several possible answers. Below, we describe computational models and laboratory studies documenting the ways by which allophonic treatment could come to be distinct from phonemic treatment.

PHONETIC MECHANISMS

One possibility for learning the difference between allophones and phonemes is that phonological status may be partially coded in the acoustic signal. Specifically, it may be that allophonic alternates are less distant from each other than phonemic ones; this difference could ensue because speakers produce them less clearly since their listeners pay little attention to them and thus communication is not compromised by their lack of distinctiveness; or simply because speakers themselves do not hear the difference very clearly, and thus never hyperarticulate these sounds. Such a strategy appears to be a cheap and sensible one, since infants are extremely sensitive to the acoustic properties of phonemes in their input (Maye et al., 2002; McMurray and Aslin, 2005; Cristia et al., 2011).

Corpus studies confirm that phonological status is, indeed, coded in the acoustic signal. Yuan and Liberman (2011) measured the Mel-frequency cepstral coefficients (MFCCs) of nasal and oral vowels in three languages (Mandarin, Portuguese, English) and after training used a classifier to sort the vowels into either nasal or oral classes. Results revealed that classification was easier for Portuguese, a language with phonemic nasality, than in either English or Mandarin, languages in which nasality is allophonic. Thus, these data support the idea that there may be acoustic cues to the classification of either phoneme or allophone, such that phonemes are more distinct and hence more easily classified using MFCCs.

Similar findings may obtain in infant-directed speech. In recent work, Cristia et al. (2010) measured two different phonemic and allophonic contrasts in infant- and adult-directed speech in corpora of Quebec French and American English. Specifically, they explored tenseness which is phonemic in English (“bit” vs “beet”), but allophonic in Quebec French: In Quebec French tense vowels
are lax in closed syllables. They also explored vowel nasalization which is phonemic in Quebec French (“mode” vs “monde”), but allophonic in English: In English vowels are nasalized before tautosyllabic nasal consonants. After collecting corpora of both tense and lax, and nasal and oral vowel pairs in each language in phonologically controlled environments and in both infant- and adult-directed registers, they conducted acoustic measures of Euclidean distance between vowel-specific alternates (nasal/oral, tense/lax) using traditional acoustic measures of tenseness and nasality. Results revealed that in terms of acoustic distance the tense/lax pairs of vowels were closer in the allophonic language than in the phonemic one regardless of the specific vowels explored. Nasality, on the other hand, was equally marked in both the phonemic (French) and the allophonic (English) language. While it may be the case that this uneveness was found because nasality is simply more difficult to measure acoustically than tenseness, if we take this data at surface value it appears that the phonemic vs allophonic distinction is better marked in some areas of acoustic space than others.

Although some information on phonemic status is clearly present in the signal, corpora studies cannot reveal whether the infant learner actually uses this acoustic information about the “closeness” of sounds in her phonological processing. Further work is necessary to answer this question.

While the argument of phonetic similarity is convincing for some cases of allophony, it is unlikely that it could explain perceptual desensitization for all sounds that adults treat as allophones. An intuitive case in point is that of /t/ allophones in English varieties, which can sometimes (albeit rarely) be realized as glottal stops. There is a priori no reason to imagine that [t] and [ʔ] are similar; and certainly not more similar than [k] and [ʔ] (that is, if [ʔ] has to be the allophone of some sound, phonetically it is much closer to /k/ than /t/). In view of such arguments, researchers have also explored other mechanisms, to which we turn.

**DISTRIBUTIONAL MECHANISMS**

An additional possibility is that infants use distributional cues, meaning the context in which a phone occurs, to discern between allophones and phonemes. For example, in English aspirated /t/ and unaspirated /t/ do not occur in the same location, so complementary distribution can effectively be used as a key to the allophonic categorization of sounds in classical phonemic versus allophonic cases. This strategy seems a sensible one since evidence suggests that young babies may be sensitive to distributions of syllables (e.g., Saffran et al., 1996) and sounds (e.g., Chambers et al., 2003; Seidl and Buckley, 2005; Cristià and Seidl, 2008; Seidl et al., 2009).

These distributional mechanisms have received support from a recent artificial grammar learning study. White et al. (2008) explored the effects on infants’ perception of exposure to an artificial grammar that could be described as having morphophonologically conditioned allophony. Specifically, they familiarized 8- and 12-month-old infants with a grammar containing “determiners” followed by “content” words in which voicing of the initial C of the content word alternated as a function of the voicing of the final segment of the function word, but only with consonants of certain manners. Note that this represents a slightly different sort of allophony than the sorts discussed above, since the “complementary distribution” did not apply within the “content” words, but it was nonetheless still predictable. While 8-month-olds were able to learn these patterns, only 12-month-olds seemed to have grouped the alternate variants into a single functional category.

In addition, computational models also provide some support to the complementary distribution strategy. Peperkamp et al. (2006) investigated the performance of a model that categorized sounds in complementary distribution as allophones, and sounds with overlapping distributions as phonemes. This algorithm was tested on both an artificial language as well as a simplified corpus of phonetically transcribed French. While the algorithm did well in correctly tagging allophones in the artificial grammar, its performance was more error-prone in the French language corpus. Specifically, it over-generated, generating allophonic alternatives that were not actually present in French. Errors of this kind were reduced to a certain extent if phonetic proximity was also taken into account.

Peperkamp et al. (2006) also suggest that these errors occur because of the presence of many near-complementary distributions, as mentioned above. Specifically, it is the cases that exist along the continuum between allophones and phonemes, but not at the edges of this continuum, which are difficult for the algorithm to correctly classify. These may be problematic to all learning algorithms of this kind (and, though evidence does not yet support this, to infants as well!). However, since near-complementary distributions are present in natural languages and there is no clear cut-off point along the continuum that has been found, it may be that until we discover how humans process these cases along the continuum we will not be able to create algorithms to do so.

In concert, experimental and modeling results support the contribution of distributional information for learning of certain cases of allophony. They also underline that distributional information alone is not sufficient, but must be packaged together with acoustic similarity. This is not a limitation, as it is likely that multiple mechanisms work in concert for the discovery of phonological status.

**LEXICAL MECHANISMS**

The most informed, or high-level, source of information for phonological status involves semantic knowledge. Jakobson (1966) proposed that children use semantic cues, essentially using minimal pairs to discern which phonemes are crucial to the input language and which are not. Thus, a child might hear palatalization in English before [j,i,e]. Thus, she will hear at least two different alternate pronunciations of the word hit. Specifically, she will hear hi[ʃ] you for “hit you,” but also hear hi[t] him for “hit him.” Both of the utterances will be uttered on occasions where hitting takes place. On a lexical account, the child would decide that these two instances of hit must map to the same underlying structure, /hit/. In addition, the child will be at the same time learning which sounds are phonemes by calculating minimal pairs. Thus, the child will learn that /s/ and /h/ are distinct phonemes of English because sit and hit map onto different semantic representations. Indeed, Yeung and Werker (2009) experimentally demonstrated that infants regain attention to a non-native contrast after seeing
the members of the contrast paired with different visual referents. These two processes, one of semantic overlap and one of semantic distinction, may occur together and drive children's developing phonological representations. In a certain sense we can rule out the strong version of this hypothesis as the sole method of learning given that infants at 11 months in Seidl et al. (2009) treated allophones as distinct from phonemes. Specifically, because infants at 11 months (and likely even older: Dietrich et al., 2007) do not have many minimal pairs (Caselli et al., 1995) it seems unlikely that they can use lexical cues as the sole driving factor in their phonological category learning.

Thus, the old-style lexical hypothesis seems not to hold much promise. However, a new version of lexical bootstrapping has emerged in recent years. This work is based on the finding that minimal pairs can be insufficient for the learner to maintain a phonological distinction, and that near-minimal pairs are more useful for deciding on phonemic dimensions. Thiessen (2007) documents that 14-month-olds presented with a perfect minimal pair based on stop voicing (such as taw-daw) fail to discriminate two syllables differing along that feature, whereas toddlers exposed to near-minimal pairs (such as tawbow and dawgo) have an easier time. Swingley (2009) goes further to propose that infants could use commonalities in the pronunciation across otherwise completely different forms (such as the first vowel in yellow and better, something one could describe as "maximal pairs") to extract sound categories, and argues that this new type of lexical bootstrapping could make a considerable contribution to infants' phonological acquisition. Swingley and collaborators have recently bolstered this case by reporting that 6-month-olds have referential knowledge of several words (Bergelson and Swingley, 2012), such that their lexicon could be slightly larger than previously thought (Tincoff and Jusczyk, 1999). Moreover, corpora analyses showed that infant-directed speech offers few true minimal pairs, but rich maximal pairs structure, which an informed machine learner can profit from to learn about the phonemes of her input language (Swingley, 2009). We expect that a similar training study with infants is underway, which would constitute the final pre-requisite for this view of lexical bootstrapping. These new versions of lexical bootstrapping assume that infants can use semantic information to pull apart phonological categories. It should follow, then, that in the absence of such separating forces, infants could collapse allophonic sounds. More specifically, if maximal pairs are necessary to establish sounds as contrastive then the absence of such pairs may aid in establishing similarities between structures and assigning phonological alternations/allophonic relationships. Thus, this same mechanism might help the toddler establish that the I in Id and Im map onto the same representation. To our knowledge, the latter argument has not been made by proponents of lexical bootstrapping of phonology, but we foresee such a theoretical development within that promising line of work.

A second strain of models of phonological acquisition does not assume rich semantic representations to separate the sounds, but proposes that infants hold a pseudo-lexicon, a dictionary of frequently encountered wordforms (Martin et al., in press). In this proposal, wordform minimal pairs are used to detect allophones, such that if the child's lexicon contains two (long) sequences of sounds that are identical except for one sound, then the two sounds that differ across the two stored sequences should be considered allophones of the same phoneme. Using such an algorithm, phones could be classified as allophones and phonemes with a much greater accuracy than with other algorithms using only distributional information, or a combination of distributional information and acoustics (detailed in the Distributional mechanisms section). A pre-requisite for this type of lexical bootstrapping is that the child has a proto-lexicon, a wordform repository. Recent experimental work corroborates this: 11-month-olds showed no preference between sequences of phones that were frequent in their input, but which did not form real words, and actual real words (Ngon et al., in press). In contrast, they do prefer frequent words over infrequent words (Hallé and de Boysson Bardies, 1994), and frequent sequences over infrequent wordforms, even when phonotactics had been controlled for (Ngon et al., in press). The next step in the exploration of this potential explanation for phonological acquisition involves showing that infants use minimal wordform pairs to collapse across the distinction, rather than separate it. If this prediction holds, it would demonstrate that minimal wordform pairs and true, lexical minimal pairs do not operate in the same fashion at all.

A variant of the latter hypothesis could be proposed where long-term storage and the assumption of different mechanisms governing wordform and lexical minimal pairs are unnecessary. It is well known that infant-directed speech abounds in repetition, with a much greater narrowing of focus than adult-directed speech (McRoberts et al., 2009). In other words, it appears that infant-directed speech exaggerates “burstiness” (Baayen, 2001), the tendency for lexical items to recur within the same conversational interaction, in a way that could influence phonological acquisition (Skoruppa et al., 2012). A smart learner may be able to use variation across two wordforms experienced in close succession to derive probabilities of non-contrastiveness. For example, if the child hears “dad,” “da[dl],” “da[rl]” in the same conversational interaction, she may be able to store that [d] and [r] could be variants of the same phoneme. The latter extension has not yet been espoused by modelers, but we expect it may be just around the corner. The predictions from this hypothesis could also be easily tested using an artificial grammar design.

Whereas the combination of acoustics and distributional cues seemed to gain the learner-model quite a bit, some work suggests that a learner-model combining distributional and lexical mechanisms, or all three together, may only be subtly improved (Boruta, 2011). It is of theoretical and empirical interest to thoroughly investigate the effects and interactions emerging from the integration of all 3 types of mechanisms in the future.

**Implications**

Collectively, this work suggests that multiple mechanisms, likely including the computation of complementary distribution and the calculation of phonetic similarity, operate in concert to guide infants toward their functional interpretation of sounds that are present in the input, yet not contrastive. This review also bears on the more general question of how infants cope with phonetic variability that is not lexically meaningful such as variation between talkers’ voices and accents. Interestingly, infants become resilient...
to talker and accent changes also toward the end of the first year of life (Houston and Jusczyk, 2000; Schmale et al., 2010). Future work should investigate whether this similarity is merely superficial, or whether it is indicative of a perceptual reorganization allowing toddlers to recognize wordforms in the presence of lexically irrelevant variation. To answer this question, research should focus on how infants cope with deviations from canonical productions and how predictable those productions are. Moreover, the question of allophones is a categorical one, but many sources of variance are gradient and future work should explore whether these different kinds of variation are more or less learnable since it may be that gradient changes to the acoustic character of a sound are more variable.

A second consideration relates to the nuances in the concepts of phonemes and allophones laid out above, and predictions that can be stated on their learnability. Recent artificial grammar learning work suggests that infants tend to attend more to regular, neither entirely predictable nor entirely unpredictable, patterns (Gerken et al., 2011). In the domain of allophonic learning this might translate to different attention being allotted to patterns that are halfway between allophones and phonemes because of their very irregularity, a matter that could be investigated by assessing infants’ acquisition of different types of phonemes/allophones. Additionally, one could imagine that for infants the areas of the grammar in which the irregularity resides may be very important. For example, if the irregularity is lexically or morphologically based the language learning infant may not be immediately aware of it, and so would initially treat the pattern as if it were regular.

Additionally, differential processing of allophones and phonemes could inform translational research. For example, some work suggests that inappropriate learning of sounds in terms of these sound classes (e.g., perceiving equally well different phonemes and different allophonic alternates) correlates with reading ability and differs between normally developing and dyslexic children (Serniclaes et al., 2004). If we can pinpoint these differences in early development it may be possible to intervene while these infants are still at a very plastic stage of development.

Thus, longitudinal studies exploring allophonic and phonemic processing may well contribute to early intervention at some point in the future.

Although we have steered clear of production in this review, it is certainly the case that accurately representing sounds as mapping onto distinct phonemes or the same phone may relate to production, since the target phonology for production will require the child to use the underlying sound in different ways in different environments. All signs indicate that this is a process that occurs quite early in development (Fikkert and Freitas, 2006). Still it is unclear to what degree the continuum between allophones and phonemes relates to production of those categories. We leave that question for a future review, but mention here that it is crucial to unite these two processes within the infant in order to truly understand the course of infant development.

It remains unclear how infants might make use of “phonetic similarity” in discovering allophones and distinguishing them from phonemes. For example, all vowels are more similar when compared with consonants, yet even young infants do not appear to have difficulty in distinguishing one vowel from another. It may be crucial to discern how acoustic similarity is judged vis-a-vis the infant. It is possible that lexical factors may also play a role in infant learning of phonological categories in a greater way than has been shown in learning models (Swingley, 2009).

Finally, it is clear that allophones may be relevant not just to phonological learning, but also to syntactic learning since allophonic alternates may mark phrasal edges (Selkirk, 1984; Nespor and Vogel, 1986; Seidl, 2000) and this marking may help infants to learn their syntactic structure if they are attentive to these edges (Nespor et al., 1996; Christophe et al., 1998). For example, if there is strengthening of contact at domain edges (Keating et al., 2003) or specific phonological processes at domain edges as mentioned above, e.g., a greater degree of aspiration or longer linguo-palatal contact the higher up you go in the prosodic hierarchy, then if infants are aware of the prosodic cues that they use for syntactic bootstrapping, this knowledge should inform or at least interact with their acquisition of the knowledge of allophones.

REFERENCES
Baayen, R. H. (2001). Word Frequency Distributions. Dordrecht: Kluwer Academic Publishers.
Bazell, C. E. (1954). "The choice of criteria in structural linguistics," in Linguistics Today eds A. Martinet and U. Weinreich (New York: Linguistic Circle of New York).
Bergelson, E., and Swingley, D. (2012). At 6 to 9 months, human infants know the meanings of many common nouns. Proc. Natl. Acad. Sci. U.S.A. 109, 3253–3258.
Bertoncini, J., Bibeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., and Mehler, J. (1988). An investigation of young infants’ perceptual representations of speech sounds. J. Exp. Psychol. Gen. 117, 21–33.

Seidl and Cristia
Infants’ learning of phonological status

Frontiers in Psychology | Language Sciences
November 2012 | Volume 3 | Article 448 | 8

Analysis of Signs, eds H. van der Hulst and R. Channon (Nijmegen: Ishara Press), 305–60.
Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sander, L., et al. (1995). A cross-linguistic study of early lexical development. Cogn. Dev. 10, 159–199.
Chambers, K., Onishi, K., and Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. Cognition 87, B69–B77.
Christophe, A., Guasti, M. T., Nespor, M., and van Ooyen, B. (1998). Prosodic structure and syntactic acquisition: the case of the head-complement parameter. Dev. Sci. 6, 213–222.
Clements, G. N. (2001). "Representational economy in constraint-based phonology," in Distinctive Feature Theory, ed. T. A. Hall (Berlin: Mouton de Gruyter), 71–146.
Cristia, A., McGuire, G., Seidl, A., and Francis, A. (2011). Effects of the distribution of cues on infants’ perception of speech sounds. J. Phon. 39, 388–402.
Cristià, A., and Seidl, A. (2008). Is infants’ learning of sound patterns constrained by phonological features? Lang. Learn. Dev. 4, 203–227.
Cristia, A., Seidl, A., and Onishi, K. H. (2010). Indices acoustiques de phonémicité et d’allophonie dans la parole adressée aux enfants. Actes des Journées d’Étude sur la Parole 28, 277–280.
Infants’ learning of phonological status

Dahan, D., Drucker, S., and Scarborough, R. (2008). Talker adaptation in speech perception: adjusting the signal or the representation? Cognition 108, 710–718.

Dietrich, C., Swingley, D., and Werker, J. (2007). Native language governs interpretation of salient speech sound differences at 18 months. Proc. Natl. Acad. Sci. U.S.A. 104, 454–464.

Fikkert, P., and Freitas, M. J. (2006). Allophony and allomorphy cue phonological development: evidence from the European Portuguese vowel system. J. Catal. Ling. 5, 83–108.

Fischer, J. L. (1958). Social influence in the choice of a linguistic variant. Word 14, 47–56.

Gerken, L., Balcomb, F., and Minton, J. (2011). Infants avoid ‘labouring in vain’ by attending more to learnable than unlearnable linguistic patterns. Dev. Sci. 14, 972–979.

Haucard, V., Walter, M. A., and Marantz, A. (2007). The effects of inventory on vowel perception in French and Spanish: an MEG study. Brain Lang. 100, 295–308.

Hall, K. C. (2009). A Probabilistic Model of Phonological Relationships from Contrast to Allomorphy. PhD thesis, The Ohio State University, Columbus.

Halle, M. (1964). “On the bases of phonology,” in The structure of language: readings in the philosophy of language, eds J. A. Fodor and J. J. Katz (Englewood Cliffs, NJ: Prentice-Hall), 324–333.

Hallé, P., and de Boysson-Bardies, B. (1994). Emergence of an early lexicon: infants’ recognition of words. Infant. Behav. Dev. 17, 119–129.

Hohne, E. A., and Jusczyk, P. W. (1994). Two-month-old infants’ sensitivity to allophonic differences. Percept. Psychophys. 56, 613–623.

Houtson, D., and Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. J. Exp. Psychol. Hum. Percept. Perform. 26, 1570–1582.

Jakobson, R. (1966). “Beitrag zur allgemeinen,” in Readings in Linguistics II (Chicago: University of Chicago Press), 51–89.

Johnson, K., and Babel, M. (2010). On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. J. Phon. 38, 127–136.

Jusczyk, P. W., Hohne, E., and Bau- man, A. (1999). Infants’ sensitivity to allophonic cues for word segmentation. Percept. Psychophys. 61, 1465–1476.

Kazanina, N., Phillips, C., and Idsardi, W. (2006). The influence of meaning on the perception of speech sound contrasts. Proc. Natl. Acad. Sci. U.S.A. 103, 11183–11186.

Keating, P., Cho, T., Fougeron, C., and Hsu, C.-S. (2003). “Domain-initial articulatory strengthening in four languages,” in Papers in Laboratory Phonology VI, eds J. Local, R. Ogden, and R. Temple (Cambridge: Cambridge University Press), 143–161.

Kenstowicz, M. (1994). Phonology in Generative Grammar. Cambridge, MA: Blackwell.

Koudaurova, M., and Francis, A. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. J. Acoust. Soc. Am. 124, 3959–3971.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). Philos. Trans. R. Soc. Lond. B Biol. Sci. 363, 979–1000.

Kuhl, P. K., Stevens, E., Hayashi, A., Kiritan, S., Kiritan, S., and Iverson, P. (2006). Infants show a facilitatory effect for native language phonetic perception between 6 and 12 months. Dev. Sci. 9, 13–21.

Ladd, R. (2006). “Distinctive phonés in surface representation,” in Laboratory Phonology Vol. 8, eds L. M. Goldstein, D. H. Whalen, and C. T. Best (Mouton de Gruyter), 3–26.

Martin, A., Peperkamp, S., and Dupoux, E. (in press). Learning phonemes with a proto-lexicon. Cogn. Sci. 40, 82–96.

Maye, I., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can effect phonetic discrimination. Cognition 82, B101–B111.

McLaughlin, C. T., Luce, P. A., and Charles-Luce, J. (2003). Representation of lexical form. J. Exp. Psychol. Learn. Mem. Cogn. 4, 129–134.

McMurray, R., and Aslin, D. (2005). Infants are sensitive to within-category variation in speech perception. Cognition 95, B15–B26.

McRoberts, G. W., McDonough, C., and Lakusta, L. (2009). The role of verbal repetition in the development of infant speech preferences from 4 to 14 months of age. Infancy 14, 162–194.

Narayan, C. (2006). Acoustic-Perceptual Salience and Developmental Speech Perception. PhD thesis, University of Michigan.

Nespoulous, M., Guasti, M., and Christophe, A. (1996). “Selecting word order: The Rhymtic Activation Principle in Infonology ed. U. Kleinheinz (Berlin: Academ Verlag), 1–26.

Nespoulous, M., and Vogel, I. (1986). Prasodie Phonology. Dordrecht: Foris.

Ngon, C., Martin, A., Dupoux, E., Cabrol, D., Dutat, M., and Peperkamp, S. (in press). (Non)words, (non)words, (non)words: evidence for a protolinguistic stage of the first year of life. Dev. Sci.

Pegg, J., and Werker, J. F. (1997). Adult and infant perception of two English phones. J. Acoust. Soc. Am. 101, 3742–3753.

Peperkamp, S., LeCâvez, R., Nadal, J. P., and Dupoux, E. (2006). The acquisition of allophonic rules: statistical learning with linguistic constraints. Cognition 101, B31–B41.

Peperkamp, S., Pettinato, M., and Dupoux, E. (2003). “Allophonic variation and the acquisition of phoneme categories,” in Proceedings of the 27th Annual Boston University Conference on Language Development, eds B. Reachley, A. Brown, and F. Conlin (Bostom: Cascadilla Press), 650–661.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. Lang. Speech 3, 115–154.

Pike, K. (1947). Phonemics. The University of Michigan Press, Ann Arbor.

Polka, L., Colantoni, C., and Sundara, M. (2001). A cross-language comparison of /id/-/it/ perception: evidence for a new developmental pattern. J. Acoust. Soc. Am. 109, 2190–2201.

Peperkamp, S., and Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. J. Exp. Psychol. Hum. Percept. Perform. 20, 421–435.

Saffran, J., Aslin, R., and Newport, E. (1996). Statistical learning by 8-month-old infants. Science 274, 1926–1929.

Schmole, R., Cristia, A., Seidl, A., and Johnson, E. K. (2010). Infants’ word segmentation across dialects. Infancy 15, 650–662.

Scobbie, J., Turk, A., and Hewlett, N. (1999). “ Morphemes, phonetics and lexical items: the case of the Scots vowel length rule,” in Proceedings of the XVIIIth International Congress of Phonetic Sciences, San Francisco, 1617–1620.

Scobbie, J. and Stuart-Smith, J. (2008). “Quasi-phonemic contrast and the indeterminacy of the segmental inventory: Examples from Scottish English,” in Contrast in phonology: Perception and Acquisition, eds P. Avery, B. E. Dresher, and K. Rice (Mouton: Berlin).

Seidl, A. (2000). Minimal Indirect Reference: A Theory of the Syntax-Phonology Interface. New York: Routledge.

Seidl, A., and Buckley, E. (2005). On the learnability of arbitrary phonological rules. Lang. Learn. Dev. 1, 289–316.

Seidl, A., Cristia, A., Onishi, K., and Bernard, A. (2009). Allophonic and phonemic contrasts in infants’ learning of sound patterns. Lang. Learn. Dev. 5, 191–202.

Selkirk, E. (1984). Phonology and Syntax: The Relation between Sound and Structure. MIT Press, Cambridge, MA.

Serniclaes, W., Heghe, S. V., Moutou, P., Carre, R., and Sprger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. J. Exp. Child. Psychol. 87, 336–361.

Shea, C., and Curtin, S. (2011). Experience, representations and the production of second language allophones. Cognition 27, 229–250.

Skoruppa, K., Mani, N., and Peperkamp, S. (2012). Toddlers’ processing of phonological alternations: Early compensation for assimilation in English and French. Dev. Sci.

Swingley, D. (2009). Contributions of infant word learning to language development. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364, 3617–3622.

Thiessen, E. (2007). The effect of distributional information on children’s use of phonemic contrasts. J. Mem. Lang. 56, 16–34.

Tincoff, R., and Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. Psychol. Sci. 10, 172–175.
Trubetzkoy, N. S. (1939/1969). Principles of Phonology, eds A. Christiane and M. Baltaxe, Trans (Berkeley: University of California Press).

Werker, J. F., and Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorga-nization during the first year of life. Infant Behav. Dev. 7, 49–63.

Whalen, D., Best, C., and Irwin, J. (1997). Lexical effects in the percep-tion and production of American English /p/ allophones. J. Phon. 25, 501–528.

White, K. S., Peperkamp, S., Kirk, C., and Morgan, J. (2008). Rapid acquisition of phonological alter-nations by infants. Cognition 107, 238–265.

Yeung, H., and Werker, J. (2009). Learning words’ sounds before learn-ing how words sound: 9-month-old infants use distinct objects as cues to categorize speech information. Cognition 113, 234–243.

Yuan, J. and Liberman, M. (2011). “Automatic measurement and com-parison and vowel nasalization across languages,” in Proceedings of ICPH$^5$ XVII (Hong Kong: ICPHS), 2244–2247.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any com-mercial or financial relationships that could be construed as a potential con-{}