American Psychological Association

Developmental Psychology

0012-1649/19/$12.00

2019, Vol. 55, No. 5, 934–950
http://dx.doi.org/10.1037/dev0000699

Associative Word Learning in Infancy: A Meta-Analysis of the Switch Task

Angeline Sin Mei Tsui
University of Ottawa

Krista Byers-Heinlein
Concordia University

Christopher T. Fennell
University of Ottawa

Associative word learning, the ability to pair a concept to a word, is an essential mechanism for early language development. One common method by which researchers measure this ability is the Switch task (Werker, Cohen, Lloyd, Casasola, & Stager, 1998), wherein infants are habituated to 2 word-object pairings and then tested on their ability to notice a switch in those pairings. In this comprehensive meta-analysis, we summarized 141 Switch task studies involving 2,723 infants of 12 to 20 months to estimate an average effect size for the task (random-effect model) and to explore how key experimental factors affect infants' performance (fixed-effect model). The average effect size was low to moderate in size, Cohen's $d = 0.32$. The use of language-typical and dissimilar-sounding words as well as the presence of additional facilitative cues aided performance, particularly for younger infants. Infants learning 2 languages at home outperformed those learning 1, indicating a bilingual advantage in learning word-object associations. Together, these findings support the Processing Rich Information from Multidimensional Interactive Representations (PRIMIR) theoretical framework of infant speech perception and word learning (e.g., Werker & Curtin, 2005), but invite further theoretical work to account for the observed bilingual advantage. Lastly, some of our analyses raised the possibility of questionable research practices in this literature. Therefore, we conclude with suggestions (e.g., preregistration, transparent data peeking, and alternate statistical approaches) for how to address this important issue.

Keywords: Switch task, infancy, word learning, speech perception, meta-analysis

Meta-analysis, a statistical method used to summarize individual studies addressing the same research question, has been widely used in education, psychology, and medical research. Yet, to date, only a handful of such analyses have been conducted in the field of early language acquisition, despite the utility of meta-analysis to confirm and sometimes challenge established developmental theories. A confirmatory example is a meta-analysis by Tsuji and Cristià (2014), which substantiated infants' generally recognized developmental patterns of perceptual attunement to native vowels, and declines in sensitivity to nonnative vowels. However, a meta-analysis on infant word segmentation by Bergmann and Cristia (2016) challenged accepted hypotheses by showing that infants of all ages had a familiarity preference (i.e., looked longer to a familiar than a novel stimulus)—a surprising result given widely accepted models of infant information processing that posit a developmental move from familiarity to novelty preferences (e.g., Hunter & Ames, 1988). Similarly, a meta-analysis by Rabagliati, Ferguson, and Lew-Williams (2019) found that infants' abstract rule learning is domain-general, challenging the proposal that this ability is revealed only using speech stimuli (e.g., Marcus, Fernandes, & Johnson, 2007). Another powerful use of meta-analyses is to estimate average effect sizes for key experimental tasks, which can be used to plan adequately powered future studies (Bergmann et al., 2018; Wellman, Cross, & Watson, 2001). This is particularly important in the context of the replication crisis in the scientific community, including in the psychological sciences (Braver, Thoemmes, & Rosenthal, 2014; Open Science Collaboration, 2015).

In consideration of the established utility of this method, we present a meta-analysis aggregating 141 studies that have used a

This article was published Online First February 7, 2019.

Angeline Sin Mei Tsui, School of Psychology, University of Ottawa; Krista Byers-Heinlein, Department of Psychology, Concordia University; Christopher T. Fennell, School of Psychology, University of Ottawa.

This work was funded by a Social Sciences and Humanities Research Council of Canada Insight Grant (435-2015-1974) to Christopher T. Fennell and a Natural Sciences and Engineering Research Council of Canada Discovery Grant (402470-2011) to Krista Byers-Heinlein. Christopher T. Fennell holds the Official Languages and Bilingualism Institute Research Chair in Language Learning and Acquisition at the University of Ottawa; Krista Byers-Heinlein holds the Concordia University Research Chair in Bilingualism.

We thank Maxine Berthiaume and Caitlin Atkinson for their assistance of the data coding. We also want to thank the researchers who kindly tracked down their data for us and those who graciously provided us with unpublished findings.

Correspondence concerning this article should be addressed to Angeline Sin Mei Tsui, School of Psychology, University of Ottawa, Ottawa, Canada K1N 6N5. E-mail: atsui029@uottawa.ca

common laboratory task called the Switch task to measure infants' ability to associate words with objects. Word-object associations are a key component of the word learning process and, therefore, a fundamental aspect of language acquisition. Our meta-analysis aims to (a) provide effect size estimates for future Switch task studies; (b) investigate factors contributing to infants' successful associative word learning and consider the implications for theories of language acquisition; and (c) determine if evidence exists for a bias in the magnitude of the effects reported in this literature. Before describing the meta-analysis itself, we provide a short overview of associative word learning and the Switch task, highlighting their importance in infant language development.

Associative Word Learning and the Switch Task

Words are the basic units of meaning for effective communication, and the ability to learn new words develops rapidly in infancy. Around 6 months of age, infants show comprehension of some everyday words such as *eyes* and *banana* (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 2012), and by age 12 months they comprehend dozens of words and can even produce a few (Fenson et al., 1994). This rapid development is even more impressive when considering that word learning is a multifaceted task. Consider an infant who hears a parent say, "Look at the *ball.*" At minimum, the infant needs to: extract word form *ball* from the fluent speech stream, hold the word form in mind, identify the object *ball* among all the other toys as the word's referent, and associate the word form and its referent. In this article, we focus on studies that test the latter crucial component of this word learning process: associative word learning.

Associative word learning refers to the mechanism for linking a word to a referent. Some researchers suggest that perceptually based associative word learning is the main manner in which infants build their vocabulary (e.g., Sloutsky & Robinson, 2008; Smith, Jones, Yoshida, & Colunga, 2003). Others researchers have argued that, in addition to the perceptual information necessary for associative word learning, infants also leverage conceptual information as they learn new words (e.g., Booth, Waxman, & Huang, 2005; Waxman & Gelman, 2009). Nevertheless, all theorists acknowledge that associative word learning is an essential underlying mechanism in the process of early word learning. One of the more common ways to test this fundamental skill is via a looking-based laboratory paradigm called the Switch task (Werker et al., 1998).

In the standard version of the Switch task, infants are habituated to two word-object pairings (Object A paired with Word A; Object B paired with Word B). Typically, both the words and the objects are novel to the infant. Across multiple trials, the two word-object pairings are presented until infants' attention declines to a preset criterion. After this habituation, infants experience two test trials: a Same trial involving one of the original word-object pairings (e.g., Object A - Word A) and a Switch trial that entails a mismatched word-object pairing (e.g., Object A - Word B). If infants learned the original word-object pairings, they should detect the violation of the word-referent link during the Switch trial and look longer during the Switch trial than during the Same trial. The simplicity of the Switch task and its focus on novel word-object pairings has led to it becoming a common method to test infants' novel associative word learning.

Here, we report a meta-analysis of Switch task studies, an approach that can make several potential contributions to the literature on early associative word learning. On the practical side, our meta-analysis will provide an estimate of the average effect size for Switch task studies, key information for determining the minimum sample size needed to adequately power such studies. It will also explore any evidence for a bias in the magnitude of effects reported in Switch task literature, and its potential origin. As researchers in the field continue to aggregate data from different experimental tasks (see MetaLab, 2016), information about the effect size of the Switch task and any potential research biases will also help researchers choose the most sensitive tasks for their research question.

On the theoretical side, our meta-analytic approach also can provide conceptual insights by examining which demographic and experimental factors (moderators) might influence infants' performance in the Switch task; thus, providing tests of theoretical assumptions of early word learning. As such, we have identified six moderators of interest in this meta-analysis, based on variables that have been systematically examined in past Switch task studies: infant age, the number of word-object pairings learned during habituation, language background (i.e., monolingual or bilingual), word typicality (i.e., whether the novel words were word-like and consistent with the native-language phonology), word similarity (i.e., whether words were similar vs. dissimilar sounding), and the presence or absence of facilitative manipulations. Some of these moderators are of particular theoretical interest. For example, understanding how and when infants learn novel minimal pairs (i.e., phonetically similar words, like "bin" and "din"), reflected in the word similarity moderator, has been a large focus of the Switch literature because of proposals concerning how phonology and lexical acquisition interact (e.g., Rost & McMurray, 2009, 2010; Thiessen, 2007; Werker, Fennell, Corcoran, & Stager, 2002). We will, therefore, apply our findings to theories that consider the intersection of phonological and lexical development. We investigated whether these moderators, and particularly interactions with age, predict infant performance in the Switch task. We will now discuss each moderator in further detail.

Infant Age

The Switch task has been used primarily from 12 to 20 months, and we ask whether associative word learning undergoes a significant development shift over infants' second year. If no major shift occurs, it would imply that this skill is well-established by 12 months. If there is an amelioration of the skill, this might reflect the effects of cumulative experience in supporting vocabulary growth in infancy and beyond (McMurray, 2007; Mitchell & McMurray, 2009). Alternatively, age might interact with other moderators rather than simply manifesting as a general improvement, a point we return to below.

Number of Word-Object Pairings

In the standard Switch task (e.g., Stager & Werker, 1997), infants experience two word-object pairings during habituation, but in some versions infants only are presented with a single pairing (e.g., Fennell, 2012). Learning two word-object pairings might be cognitively more demanding than learning a single pair, as infants have to track two objects and two word forms in both working and long-term memory. Therefore, this moderator can test the limits of infants' word-object associative skills.

Language Background

Compared with monolinguals, bilinguals receive less exposure to each of their languages and encounter more variable language input (e.g., accented speech, language switching, cognate words; for a discussion, see Byers-Heinlein & Fennell, 2014). As such, investigating bilingual infant language development helps researchers understand and test theories of infant language acquisition (Byers-Heinlein & Fennell, 2014; Fennell, Tsui, & Hudon, 2016). There is mixed evidence with regards to whether monolinguals and bilinguals differ in associative word learning skills, with some studies reporting similarities (Fennell & Byers-Heinlein, 2014; Mattock, Polka, Rvachew, & Krehm, 2010) and others reporting differences (Fennell, Byers-Heinlein, & Werker, 2007; Graf Estes & Hay, 2015; Hay, Graf Estes, Wang, & Saffran, 2015). In view of the equivocal literature, the current meta-analysis investigates whether monolingual and bilingual infants differ in their novel word learning.

Word Typicality

An important milestone in language development is infants' ability to use their knowledge of native language properties when learning words. During the first year of life, infants exhibit increasing mastery of the specifics of their native languages, including preferential processing of native language sounds (i.e., phonology; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984) and sound sequences (i.e., phonotactics; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Jusczyk, Luce, & Charles-Luce, 1994). Based on this, infants should selectively accept only object labels that conform to native-language properties in associative word learning. Indeed, several Switch task studies have tested infants' willingness to learn illegal or ill-formed words, such as those containing nonnative phonemic information (Hay et al., 2015; May & Werker, 2014), those with low-probability or illegal phonotactics (Graf Estes, 2014; Graf Estes, Edwards, & Saffran, 2011; Mackenzie, Graham, Curtin, & Archer, 2014), and words that denote grammatical functions and not object reference (Mackenzie, Curtin, & Graham, 2012a). In general, if infants show such learning constraints for word object associations, we will see a main effect of this moderator in favor of typicality, or an interaction with age if infants require substantial native-language experience for word typicality to affect their word-learning preferences.

Word Similarity

Many Switch task studies have focused on whether minimal pairs are harder to learn than dissimilar-sounding words (e.g., Stager & Werker, 1997), similar to the literature on infants' detection of mispronunciations in familiar words (e.g., Swingley & Aslin, 2000). This interest was driven by an apparent paradox where infants refine their speech perception to efficiently process the vast majority of native-language phonemes by 12 months (Werker & Tees, 1984), yet infants of 14 months fail to utilize this well-refined phonetic sensitivity when associating minimal pairs to referents (Stager & Werker, 1997).

If similar-sounding words are truly more difficult for infants to learn than dissimilar-sounding words, this would be evinced in our meta-analysis by a main effect of word similarity. However, some developmental frameworks of word learning posit that word similarity should interact with age. According to the Processing Rich Information from Multidimensional Interactive Representations framework (PRIMIR; Werker & Curtin, 2005), infants' initially encode and try to attend to all the rich information from speech input when building their lexical representations, including both lexical (i.e., phonetic details) and nonlexical information (i.e., indexical details like speaker gender). This inclusion of irrelevant information impairs their minimal pair learning. It is only as infants acquire more lexical experience with age and thus build stronger phonemic representations that they are able to direct their attention to the relevant phonetic detail in minimal pair word learning tasks. An interaction between word similarity and age in our meta-analysis would reflect such a qualitative developmental shift in word learning, supporting the proposals found in PRIMIR.

Facilitative Manipulations

As discussed above, developmental frameworks such as PRIMIR propose that infants' ability to learn similar-sounding words improves as their phonological representations undergo a qualitative developmental shift. However, other theories, such as the Resource Limitation Hypothesis (Werker & Fennell, 2004), posit that younger infants fail to learn minimal pair words because their more immature cognitive resources prevent full integration of well-established phonological detail when connecting word and referent. To adjudicate between these two positions, researchers have used manipulations aimed at facilitating infants' minimal pair word learning: low-level facilitative manipulations that typically test the incomplete phonological representations position, and high-level facilitative manipulations that typically test the Resource Limitation Hypothesis.

Low-level cues. One reason that infants might have difficulty with learning similar-sounding words is that their initial phonological representations are underspecified and holistic (Brown & Matthews, 1997; Ferguson & Farwell, 1975). Relatedly, other researchers have argued that infant speech discrimination skills are not well-developed by 12 months; therefore, the phonological representations based on these skills would be similarly underdeveloped (Galle, Apfelbaum, & McMurray, 2015). Both approaches raise the possibility that specific perceptual experience could facilitate infants' attention and encoding of the phonological differences in similar-sounding words. Giving infants richer and more variable low-level acoustic/phonetic information during learning has been proposed to enrich the representation of relevant phonological details, for example by increasing the number of different speakers who produce the minimal pairs during habituation (Rost & McMurray, 2009), or by presenting words in infant-directed speech (Graf Estes & Hurley, 2013). These proposals predict that these low-level manipulations will help younger infants, who have less detailed phonological representations, to learn similar-sounding words.

High-level cues. Contrary to the theories that posit initially underspecified phonological representations, others suggest that infants' initial phonological representations are relatively complete, rooted in the rich speech perception skills refined over the first year of life. Infants' apparent failure in learning minimal pairs in the Switch task may stem from their limited cognitive capacity and the task demands of the Switch procedure itself (e.g., attention, memory, and association), which overwhelms their ability to attend to and incorporate detailed phonological information into lexical representations (e.g., Fennell & Werker, 2003). As infants grow older, they acquire mature lexical skills and develop stronger cognitive capacities, allowing them to successfully learn similar-sounding words by directing attention to the most relevant aspect of the task: phonetic detail (Werker et al., 2002). While lexical skills and cognitive maturation aid the infant over developmental time, this proposal raises the possibility that researchers can similarly aid infants in the lab by providing high-level cues that make the word learning process itself easier, thereby reducing cognitive demands and, in turn, improving infants' ability to learn minimal pair words. An example of a high-level manipulation is providing infants with referential cues to help them understand the word-referent connections in the Switch task, such as exposing infants to familiar word-object combinations before habituation (Fennell & Waxman, 2010; May & Werker, 2014).

In summary, we expect that including either a low-level or a high-level cue facilitates infant associative word learning; thus, we predict a main effect of *facilitative manipulations.* In addition, we will examine whether the effect size differs between studies using low-level and high-level cues. This allows us to test which types of facilitative cues best bolster learning in young infants, which can inform the theoretical positions highlighted earlier. Finally, the effects of facilitative manipulations may depend on infant age, which would be seen in a statistical interaction of these two factors. For example, if older infants have strong phonological representations, their performance learning similar words in the Switch task may not be dramatically improved by addition of low-level or high-level cues. Consequently, older infants may show smaller or no facilitative effects of cues when compared with younger infants.

The current meta-analysis combined data from 141 studies to determine the average effect size of the Switch task and how the effect size varied with key experimental manipulations. This approach can provide both theoretical and methodological insights for the field of infant language development.

Method

The Research Ethics Board (REB) at Angeline Sin Mei Tsui's home institute requires no ethical review for the use of previously collected, publicly available, anonymized data, such as meta-analysis. Thus, the data in the current article, *Associative word learning in infancy: A meta-analysis of the Switch task*, were exempted from REB review at the University of Ottawa.

Identification of Relevant Studies

A multipronged approach was used to identify relevant Switch task studies. First, we used broad terms to search studies in Google Scholar. These terms included: phonolog*, infan*, Switch*/Switch task, word recognition, phon*, phonetics, child* develop*, novel words, linguistics, and vocabulary. We also conducted a Google Scholar search of all studies that cited the initial Switch task studies (i.e., Stager & Werker, 1997; Werker et al., 1998, 2002). Further, we are familiar with the Switch task literature, so we identified articles that tested infants with the Switch task based on our knowledge. Finally, we contacted researchers who had previously published Switch task studies for any unpublished data.

The above search yielded 1,629 articles and data sets. Based on abstracts and titles, we rejected 1,546 records by removing duplicates and ineligible studies (i.e., not an experimental article or did not use the Switch task). Angeline Sin Mei Tsui carefully screened the method sections of the remaining 83 records using the criteria below. Infants had to be habituated to at least one word-object pairing and be tested with at least one Same trial and one Switch trial. We excluded studies where habituation stimuli did not include a novel word-object pairing (e.g., familiar word-object pairings or nonobjects like geometric patterns), or that did not use looking time as the dependent variable (e.g., used event-related potentials). Additional inclusion criteria were that infants in the study were typically developing and not at risk for language delay. Finally, we limited our sample to studies involving infants with a mean age of 12- to 20-months, as the majority of the studies fell within this age range. Further, this age range is important for theories of infant minimal pair word-learning, which we will address in the discussion. The above procedure was consistent with the standard meta-analytic protocol: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher, Liberati, Tetzlaff, Altman & the PRISMA Group, 2009). A PRISMA flow diagram is included in Figure 1. Overall, 42 published articles and five unpublished data sets of studies conducted from 1997 to 2017 were included in the meta-analysis (see Table 1). In total, these studies reported 141 effect sizes measured across 2,723 infants.

Coding

Each study was coded only if it tested a separate and independent participant group. Published articles often include several experiments, which raises the possibility that researchers may test the same participant group across different conditions. However, this occurred for only one article in the meta-analysis. To ensure that each effect size in the meta-analysis was independent, we randomly picked and coded only one condition in that article. Angeline Sin Mei Tsui and a trained research assistant coded all variables of interest for each study with the exception of one moderator, *facilitative manipulations*, according to a coding manual written by Angeline Sin Mei Tsui, and revised by Krista Byers-Heinlein and Christopher T. Fennell (see manual here: https://osf.io/uwe8g/). The percentage of agreement between the coding of Angeline Sin Mei Tsui and the research assistant was 99%. For the moderator *facilitative manipulations*, Angeline Sin Mei Tsui and Christopher T. Fennell coded each study independently using the coding manual, with Krista Byers-Heinlein doing reliability checks on 26 studies. For this moderator, the coding agreement was assessed by Cohen's κ and the interrater reliability was 0.9. Any disagreements were discussed until consensus was reached.

1. Effect size calculation variables. These variables included: sample size, mean looking time difference between the Same and Switch trial, corresponding *SD* of differences between trials, and the paired *t*- or *F* value comparing Same-Switch trials. We

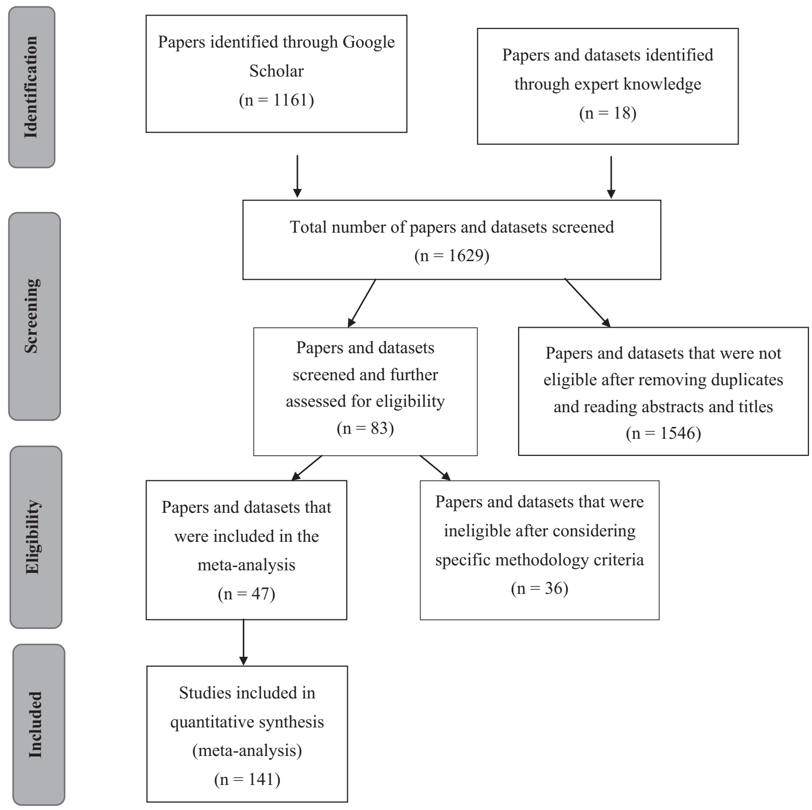


Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram that indicates the total number of articles and data sets identified, screened, and finally included in the meta-analysis. Note that multiple studies (i.e., data points in the meta-analysis) could be reported in one article or dataset. Thus, the total number of studies are larger than the number of articles and data sets included in the meta-analysis.

contacted Angeline Sin Mei Tsui of studies that did not report sufficient data for effect size calculation and 9 out of 10 replied. For studies with missing information, we estimated effect sizes using the reported *p* value, mean looking time and *SD* of Same and Switch trials, average of Same-Switch trials correlations across studies, and sample size.

2. Moderator variables. We coded each study for six moderator variables (see Table 2). We ensured that we met the recommendation of 10 or more studies per each level of categorical moderator to avoid drawing conclusions from small sets of studies (Harrell, 2001).

Effect Size Calculations

Our effect size metric was the standardized mean difference for the between-subjects Cohen's *d* (d_{IG}). This metric was chosen because it allows for straightforward comparisons with other meta-analyses. Typically, a Cohen's *d* value of 0.20 is considered as low, 0.50 as moderate, and 0.80 as a large effect (Cohen, 1988). However, as the Switch task uses a within-subjects design where looking to the Same and Switch trials is typically correlated, it is necessary to first compute a within-subject Cohen's *d* (d_{RM}) and then convert the within-subject Cohen's *d* (d_{RM}) to a between-subjects Cohen's *d* (d_{IG}) (Morris & Deshon, 2002).

For each study, we first calculated within-subject Cohen's *d* (d_{RM}) using the formula $d_{RM} = \frac{t_{RM}}{\sqrt{n}}$, where t_{RM} is the paired *t*-statistics comparing infants' average looking time to the Same test trial (M_{Same}) with their looking time to the Switch test trial (M_{Switch}), and *n* is sample size of the study (Glass, McGaw, & Smith, 1981). Next, we computed correlation *r* between the Same and Switch trials for each study, using the formula $r = \frac{SD^2 + SD^2 - SD^2}{(2)(SD_{SAME})(SD_{SWITCH})}$, where SD_{SAME} was the *SD* of the looking time to the Same trials, SD_{SWITCH} was the *SD* of the looking time to the Switch trials and SD_{DIFF} was the *SD* of the difference scores. Of the 141 effect sizes, 105 of them reported full set of summary data for the computation of correlation. The average correlation between looking time to the Same and Switch trials was $r = .39$, and we used this average correlation value to convert d_{RM} to d_{IG}. Finally, to obtain between-subjects Cohen's *d* (d_{IG}), we used the formula $d_{IG} = d_{RM}\sqrt{2(1-r)}$ (Morris & Deshon, 2002).

We conducted all analyses using the metafor package in R (Viechtbauer, 2010). To estimate the average effect size, we ran a random-effects model using DerSimonian-Laird estimator. We quantified variability in effect size across studies using Cochran's *Q* and I^2 (Borenstein, Hedges, Higgins, & Rothstein, 2009). The *Q* statistic tests whether the overall variability across studies is greater than chance. I^2 measures the proportion of the *Q* that goes beyond chance. An I^2 value of 25% is low, 50% moderate, and 75% high (Higgins, Thompson, Deeks, & Altman, 2003).

All coded data are posted online (https://osf.io/uwe8g/ and http://metalab.stanford.edu). We identified two potential outlier studies

Table 1
Studies Included in the Meta-Analysis

Studies	No of effect sizes	Infant age group (in months)	Infants' language background	Contrasts
Altvater-Mackensen and Fikkert (2010)	2	14	Dutch	faap vs. paap
Archer, Ference, and Curtin (2014)	2	14	English	dolena vs. bolena
Byers-Heinlein, Fennell, and Werker (2013)	4	12, 14	English-others (for bilinguals); English (for monolinguals)	lif vs. neem
Chan et al. (2011)	4	14, 18	English or Mandarin	keet vs. sug for English infants, bou3 vs. bou4 (tone) for Mandarin infants
Curtin (2009)	1	12	English	bedoka (lexical stress was placed either in the initial or middle syllable)
Curtin, Campbell, and Hufnagle (2012)	1	16	English	kome vs. tidu
Curtin, Fennell, and Escudero (2009)	3	15	English	deet vs. dit, deet vs. doot, dit vs. doot
Dietrich, Swingley, and Werker (2007)	6	18	Dutch or English	tam vs. taam, tam vs. tem
Fais, Werker, Cass, Leibowich, Barbosa, and Vatikiotis-Bateson (2012)	2	14	English	bin vs. din
Fennell (2006)	1	14	English	bin vs. din
Fennell (2012)	2	14	English	din vs. gin
Fennell and Byers-Heinlein (2014)	4	17	English-French (for bilinguals); English (for monolinguals)	kem vs. gem
Fennell and Tsui (2014)	2	14	English	gebin vs. gedin, goobin vs. goodin
Fennell and Waxman (2010)	3	14	English	bin vs. din
Fennell, Byers-Heinlein, and Werker (2007)	3	14, 17, 20	English-others (for bilinguals); English (for monolinguals)	bih vs. dih
Fennell, Van der Feest, and Spring (2012)	1	14	English	gin vs. din
Ference and Curtin (2015)	1	12	English	bedoka (lexical stress was placed either in the initial or middle syllable)
Fikkert (2010)	5	14, 17	Dutch	bin vs. din, bon vs. don, bon vs. din, don vs. bon, bin vs. bon
Galle, Apfelbaum, and McMurray (2015)	1	14	English	buk vs. puk
Graf Estes (2012)	3	17	English	timay, dobu, pimo, kuga
Graf Estes (2014)	3	14	English	gaffe, tove
Graf Estes and Bowen (2013)	4	19	English	moe-dike, die-moose, gay-fouth, faw-shouch
Graf Estes and Hay (2015)	2	14, 19	English-others	ku2 and ku4 (tone)
Graf Estes and Hurley (2013)	3	17	English	timay, dou, gabu, nomay
Graf Estes, Evans, Alibali, and Saffran (2007)	4	17	English	timay, dobu, nomay, gabu
Hay, Graf Estes, Wang, and Saffran (2015)	3	14, 17, 19	English	kin2 vs. kin 4 (tone)
Hay, Pelucchi, Graf Estes, and Saffran (2011)	5	17	English	bici, casa, fuga, melo (Italian words)
Mackenzie, Curtin, and Graham (2012a)	3	12	English	fep vs. wug, und vs. Iv
Mackenzie, Curtin, and Graham (2012b)	4	12	English	ptak vs. svet, plok vs. snet, mido vs. panu, hashi vs. sika
Mackenzie, Graham, and Curtin (2011)	1	12	English	fep vs. wug
Mackenzie, Graham, Curtin, and Archer (2014)	3	12	English	ptak vs. svet, keh vs. iv
Mattock, Polka, Rvachew, and Krehm (2010)	6	17	English-French (for bilinguals); English or French (for monolinguals)	bos vs. gos
May and Werker (2014)	4	14, 20	English	[!a] vs. [ǁu] (Clicks)
Pater, Stager, and Werker (2004)	1	14	English	pin vs. bin
Rost and McMurray (2009)	2	14	English	buk vs. puk
Rost and McMurray (2010)	3	14	English	buk vs puk
Singh, Fu, Tay, and Golinkoff (2018)	4	18	English-Mandarin (for bilinguals); English or Mandarin (for monolinguals)	min vs. man
Stager and Werker (1997)	3	14	English	bih vs. dih, lif vs. neem
Thiessen (2007)	5	15, 17	English	daw vs. taw
Thiessen and Yee (2010)	5	14, 15	English	yat vs. yad, dee vs. tee
Thiessen (2011)	5	15	English	da vs. dIv
Thiessen and Pavlik (2016)	2	17, 20	English	seer vs. zeer
Tsui and Fennell (2016)	2	18	English-French (for bilinguals); English (for monolinguals)	gik vs. gek

(*table continues*)

Table 1 (*continued*)

Studies	No of effect sizes	Infant age group (in months)	Infants' language background	Contrasts
Tsui, Fais, and Fennell (2018)	4	14	English	bin vs. din
Vukatana, Curtin, and Graham (2016)	4	16, 20	English	ptak vs. svet
Werker, Fennell, Corcoran, and Stager (2002)	3	14, 17, 20	English	bih vs. dih
Zhao (2014)	2	14	English	ku1 vs. ku 3 (tone), di1 vs. di3 (tone), ku2 vs. ku3 (tone), di2 vs. di3 (tone)

Note. This table includes information about the authors, year of publication, number of effect sizes for each paper/dataset, infant age group in months, native language of infants, language background (i.e., monolinguals/bilinguals), and the word contrasts used in the task. Note that individual references may contribute multiple experiments to the analyses, as shown in column 3.

(Fennell, Van der Feest, & Spring, 2012; Stager & Werker, 1997, Exp. 3) using the screening method of Viechtbauer and Cheung (2010). However, removing these studies had little effect on average effect size and did not change any findings in the moderator analyses. Therefore, we included them in the final analyses.

Testing for Effects of Moderators

We used a fixed-effects approach to examine the impact of our moderator variables, as we expected that studies at the same level of a moderator would have the same true effect (Borenstein et al., 2009; Borenstein, Hedges, Higgins, & Rothstein, 2010). We ran two different multivariate metaregression models. First, we entered all six moderators simultaneously in a metaregression model to examine the main effect of each moderator while simultaneously controlling for the other moderators. Second, we ran another metaregression model that added three interaction terms to the original model, to investigate whether *age* interacted with the following moderators: *word typicality*, *word similarity*, and *facilitative manipulations*. The residuals of each metaregression were normally distributed. The statistical significance of each moderator was tested using the Z-statistics distribution and α was set to 0.05, two-tailed. For the interaction analysis, we used a more stringent criterion for the required number of studies than for testing single moderators (minimum = 10). As the variance of an interaction is much larger than that of a main effect (Leon & Heo,

Table 2
Moderators

Variable	Categories	Description
Infant age	Continuous variable	• Most studies reported each groups' age in months (e.g., 15-month-olds) rather than a mean age (e.g., 15.4). Thus, the former value characterized each group's age.
Numbers of objects used	**1**, 2+	• Very few studies used three or more objects. As such, we coded studies using 1 versus 2 or more objects.
Language background	**Monolingual**, Bilingual	• Monolingual studies: As reported by author(s). • Bilingual studies: Infants exposed to two languages regularly from birth, heard each language >20% of the time, no systematic exposure to a third language. • No studies mixed monolingual and bilingual infants within the same sample.
Word typicality	**Typical**, Atypical	• Typical studies: Stimuli conformed to nouns in the native language of tested infants. • Atypical studies: Nonconforming stimuli, like function words, words with nonnative sounds, phonotactically illegal words, and part-words/nonwords presented in the Switch task from a preceding statistical segmentation task.
Word similarity	**Dissimilar,** Similar	• Dissimilar-sounding word studies: Tested words differed in multiple phonological elements. • Similar-sounding studies: Word pairs differed in a single phonological element (minimal pairs).
Facilitative manipulations	**Null**, Low-Level, High-Level	• Null category studies: No extra cues (i.e., basic version of original Switch task: Stager & Werker, 1997). • Low-level facilitative studies: Acoustic-phonetic information in the target words manipulated in an effort to make them more salient (e.g., placing the target contrast in a stressed syllable, increasing the voicing difference between targets) while keeping the phonological form constant; thus, making this moderator distinct from the word similarity moderator. • High-level facilitative studies: Used cues that reduced the cognitive demands of the task without manipulating the target words themselves.

Note. Reference categories are bolded.

2009), we required at least 15 studies in each group of the moderator. For example, we did not examine whether age interacts with bilingualism as there were only 12 studies examining bilingual infants.

Testing for Bias in Reported Effects

To test whether there is evidence of bias in the effects reported in the Switch task literature, we used a number of complementary approaches that will be described in more detail in the next section: a visual approach using a contour-enhanced funnel-plot diagram (Peters, Sutton, Jones, Abrams, & Rushton, 2008), and quantitative approaches using Begg and Mazumdar Rank correlation, Egger's linear regression, and p-curve analyses (Simmons, Nelson, & Simonsohn, 2011).

Results

Mean and Variance of Effect Size

The weighted mean effect size (random-effects) of the Switch task was 0.32 ($SE = 0.04$, $N = 141$, $p < .0001$), which was statistically significant and low to moderate in strength. The 95% confidence interval (CI) of effect size varied from 0.25 to 0.39. There was statistically significant heterogeneity of variance across studies (Q (140) = 354.29, $p < .0001$) and the τ^2 was 0.11. Tau-squared reflects the estimated variance of the true effect sizes across the population of studies in the meta-analysis. The I^2 was 60.48%, indicating that the heterogeneity across studies was moderate in size. The observed heterogeneity of variance supported the investigation of potential factors that moderated the observed effect sizes.

Impact of Moderators

We used a Z statistic to test all binary moderators. Only one moderator was not binary: the facilitative manipulations moderator had three levels (null, low-level, and high-level cues). The appropriate test for this moderator was an omnibus test that tests whether the regression coefficient of at least one of the target cues (either the low-level or high-level cues or both) is significantly greater than zero. Thus, the Q statistic was used to test this multilevel moderator.

Our first metaregression (Model 1) investigated main effects of each moderator. Neither age nor number of objects in habituation was associated with the effect size ($ps > 0.5$). However, bilingual infants performed better than monolingual infants ($Z = 2.16$, $B = 0.19$, $p = .03$). Moreover, infants performed better when words were dissimilar-sounding rather than similar-sounding ($Z = -2.75$, $B = -0.15$, $p = .006$), when words were typical rather than atypical ($Z = -2.94$, $B = -0.16$, $p = .003$), and when facilitative manipulations were present (QM (2) = 17.39, $p = .0002$). Finally, the effect of facilitative manipulations did not differ between studies using low-level cues and high-level cues (χ^2 (1, 133) = 0.003, $p = .96$). Here, it is important to interpret the magnitude of this increase of effect size in the context of the mean weighted effect size of 0.32 in the current meta-analysis. For example, the magnitude of the bilingual infant moderator is relatively large ($B = 0.19$), as testing bilingual infants in the task increases the effect size by 59% (i.e., 0.19/0.32 × 100).

In our second metaregression model (Model 2) exploring interactions of moderators with age, we found a statistically significant interaction between age and facilitative manipulations, indicating that these manipulations were particularly beneficial for younger infants (QM (2) = 12.21, $p = .002$; see Figure 2). There was no significant interaction between age and word similarity ($p = .102$). The interaction between age and word typicality ($p = .079$) showed a trend toward statistical significance, with word typicality affecting younger infants more than older infants. See Table 3 for details of metaregression results.

Tests for Bias in Reported Effects

One way of visually testing for publication bias is a funnel plot. In the absence of bias, plotting each study's SE against the observed effect size should yield a plot in the shape of an inverted funnel, where studies with high precision (larger sample sizes, smaller SEs) will converge toward the mean effect size, whereas studies with low precision (smaller sample sizes, higher SEs) will scatter evenly on each side of the plot. Deviations from this funnel shape can, therefore, signal bias. Visually, our funnel-plot diagram appears fairly symmetrical (Figure 3a). However, the downward sloping regression line indicates a relationship between SEs and effect sizes where smaller studies tend to report larger effect sizes. Known as the small study effect, this deviation from the expected symmetry can indicate publication bias (i.e., only small studies with significant findings submitted and accepted for publication) or outcome reporting bias (i.e., authors selectively report the best outcomes; Schwarzer, Carpenter, & Rücker, 2015). Both Begg and Mazumdar's (1994) rank correlation and Egger's linear regression (Egger, Davey Smith, Schneider, & Minder, 1997) provided statistical support for this relationship ($ps < .0001$).

We then investigated what might drive this effect. We focused on two broad possibilities: (a) the file-drawer problem and (b) questionable research practices. The file-drawer problem refers to a phenomenon where studies with significant results (i.e., relatively large effect sizes) are more likely to be published, whereas studies without significant results and with relatively low effect sizes remain unpublished (Borenstein et al., 2009). We used a contour-enhanced funnel plot (Peters et al., 2008) to see if there was evidence for suppression of nonsignificant findings in the literature (see Figure 3b). The shaded regions indicate different levels of statistical significance. A suppression of nonsignificant findings would be indicated by "missing" studies in the light and medium gray-shaded regions as these areas contain studies with p values greater than 0.05. We found no evidence for missing studies in these areas. While we acknowledge that it is unlikely that there are no Switch studies remaining in "file drawers," the issue does not appear to be particularly severe in this literature.

The second possibility is the observed publication bias emerges from the use of questionable research practices (QRP), such as including/excluding covariates or outliers in statistical analyses on the basis of post hoc results, rounding p values (John, Loewenstein, & Prelec, 2012), and conducting statistical analyses midway through experiments to determine whether to continue or stop data collection (i.e., "data peeking"; Schott, Rhemtulla, & Byers-Heinlein, 2018). Each of these QRP can be used to achieve a statistical result that is below the $p < .05$ threshold but will increase Type I error rates beyond the nominal level of .05. p values that do not follow the expected distribution across studies can be evidence that QRP influenced reported p values. Therefore,

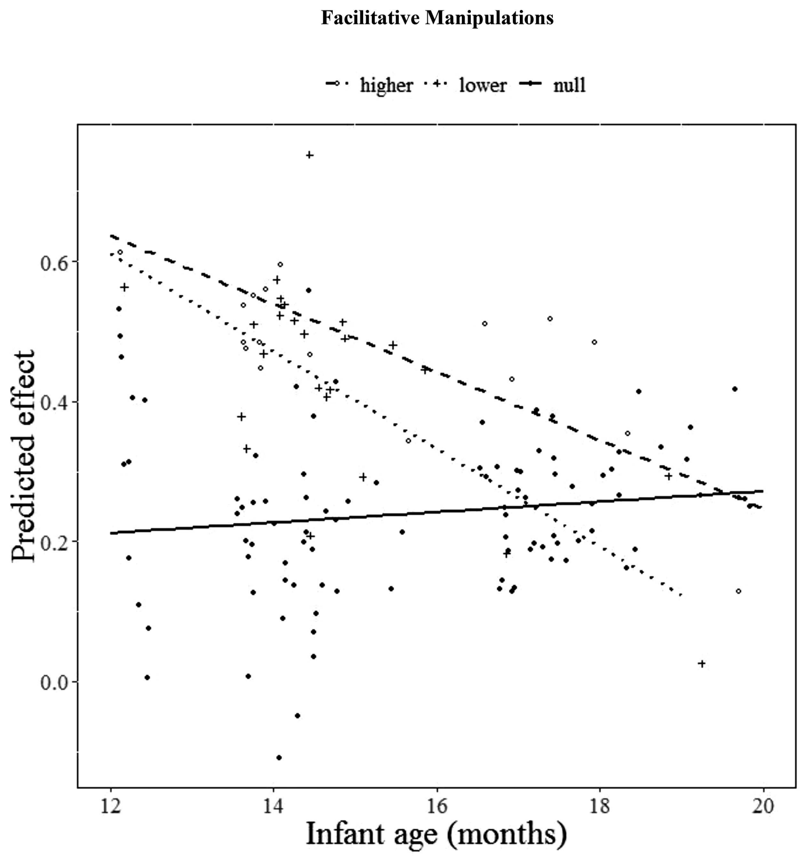


Figure 2. Predicted effect size as a function of infant age in months. Each point is a predicted effect for a unique study. In this graph, studies using higher-level cues are plotted using hollow circles, studies using lower-level cues are plotted using crosses, and the null studies are plotted using filled circles. Three different regression models are plotted here to indicate how the effect of different facilitative manipulations change when the age of infants increases. The dashed line represents the regression model for studies using higher-level cues, the dotted line represents the regression model for studies using lower-level cues and the solid line represents the regression model for the null studies.

we examined whether there was an unnatural distribution of *p* values around the significance threshold of 0.05 (Simmons et al., 2011). The *p*-curve was significantly right-skewed ($p = .0009$; see Figure 4), as expected for a literature reporting a true effect (i.e., infants can learn words in the Switch task). This suggests that selective reporting (e.g., attempting analyses with and without outliers) is unlikely to be a major issue in this literature.

We did find evidence in our data for one specific type of QRP. Only four of the 141 studies in the meta-analysis used one-tailed significance tests and all these tests would have been only "marginally" significant with the typical two-tailed test (*p*s = .06–.10). The remaining 137 studies used two-tailed tests and nearly all (135 of 137) had a clearly significant ($p < .05$) or clearly nonsignificant ($p > .10$) result. This hints that the decision of whether to conduct a one-tailed versus two-tailed-test was not decided before analysis, but instead was determined after observing the obtained *p* value, a practice which increases Type I error.

Discussion

This meta-analysis summarized data from 141 studies with a total sample size of 2723 infants to provide a quantitative review of infants' performance in the Switch task. We found a statistically significant low to moderate effect size. In other words, on average across studies, infants do look longer at the Switch trial with the word-object pairing violation than at the Same trial, but the average magnitude of this difference across studies is relatively small. The average effect size in our meta-analysis is comparable with those reported in other meta-analyses that looked at different experimental paradigms for studying infant language acquisition (e.g., Bergmann & Cristia, 2016; Tsuji & Cristià, 2014). The small learning effect size across infant studies may indicate young infants' language skills are relatively fragile and thus subject to large variation across different contexts. Indeed, we found that several significant moderators (as discussed below) can systematically explain this context variability.

Moderator Effects

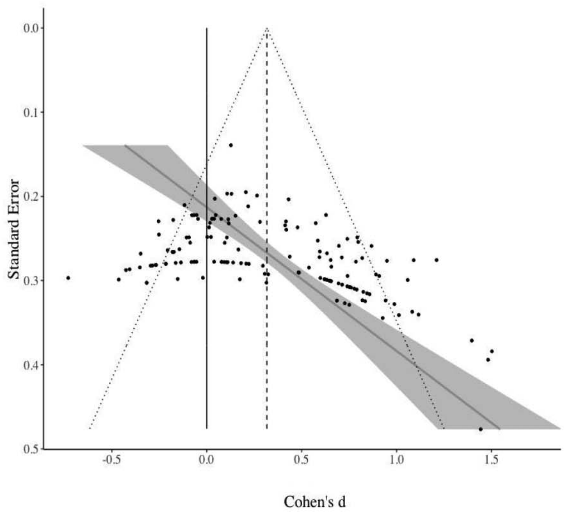
One goal of this meta-analysis was to examine how different theory-relevant moderators affected infants' performance in the Switch task. First, perhaps unexpectedly, age did not predict infants' learning in the Switch task, even when we controlled for other moderators. This result was surprising since older infants in

Table 3
Moderators of Meta-Regression Models

Moderator	Model 1			Model 2		
	B	*Z*	*p*	*B*	*Z*	*p*
Intercept	.39*	2.21	.027	.34	1.44	.151
Infant age	−.0006	−.06	.954	.002	.16	.877
Numbers of objects used in habituation	−.01	−.16	.871	−.026	−.42	.675
Language background (monolingual infants coded as 0)	.19*	2.16	.031	.19*	2.10	.035
Word typicality (typically formed words coded as 0)	−.16*	−2.94	.003	−.76*	−2.09	.037
Word similarity (dissimilar-sounding words coded as 0)	−.15*	−2.75	.006	−.78*	−2.02	.044
Low-level facilitative manipulations (null and high-level categories coded as 0)	.22*	3.25	.001	1.86*	3.00	.003
High-level facilitative manipulations (null and low-level categories coded as 0)	.22*	3.09	.002	1.50*	3.03	.002
Age × Word Typicality				.041	1.76	.079
Age × Word Similarity				.040	1.64	.102
Age × Low-level facilitative manipulations (null and high-level categories coded as 0)				−.11*	−2.66	.008
Age × High-level facilitative manipulations (null and low-level categories coded as 0)				−.081*	−2.61	.009
R^2		.09			.13	
Changes in *Q*		31.23**			44.79**	

* $p < .05$. ** $p < .0001$.

our analysis (i.e., 18 to 20 months) would typically have accumulated a substantial number of words and it has been hypothesized that associative word learning skills should improve as the early acquisition of words fuels the acquisition of new, subsequent words (McMurray, 2007; Mitchell & McMurray, 2009). Thus, we had expected improvements in infant associative word learning from 12 to 20 months (see Werker et al., 2002). One possible reason for the lack of an age effect is that associative word learning is fairly established by 12 months (e.g., Byers-Heinlein, Fennell, & Werker, 2013; Mackenzie et al., 2012a); thus, this basic ability is stable past that age. While this does not negate the many other ways that word learning develops past this age in manners not



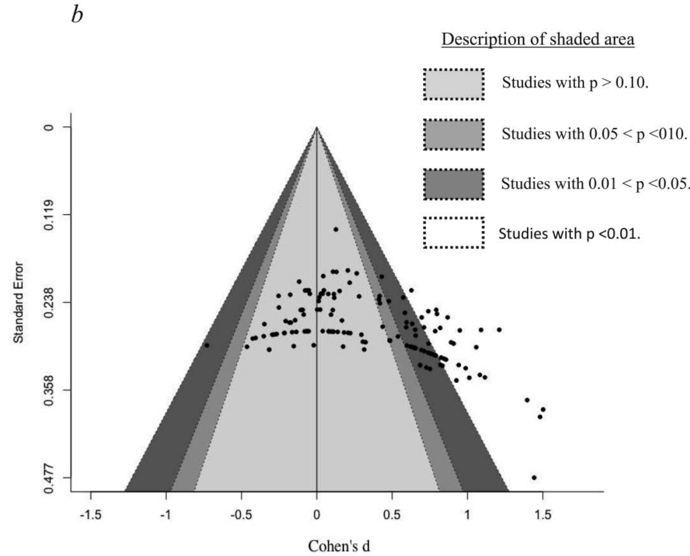


Figure 3. Funnel plots. Figure 3a depicts the funnel plot that is centered at the average effect size of the current meta-analysis. As in a typical funnel plot, the direction of the *y*-axis is negative where the *SE* increases when moving down the *y*-axis. The downward sloping regression line indicates a positive relationship between *SE* and effect sizes, suggesting that studies with smaller sample sizes tend to report larger effect sizes. The shaded area along the regression line indicates the 95% confidence interval of the regression model between *SE* and effect size. Figure 3b depicts the contour-enhanced funnel plot. Here, the plot is centered at zero as it represents no learning (i.e., infants look to the Same and Switch trials equally). Different gray areas represent studies with different *p* values. The light gray area contains studies with *p* values larger than 0.10, the medium gray area contains marginally significant studies with *p* values between 0.05 to 0.10, the dark gray area contains significant studies with *p* values between 0.01 to 0.05, and the white area contains highly significant studies with *p* values lower than 0.01. A large portion of studies lay in the light and medium gray areas, suggesting that suppression of nonsignificant results in the current Switch literature is not a major issue.

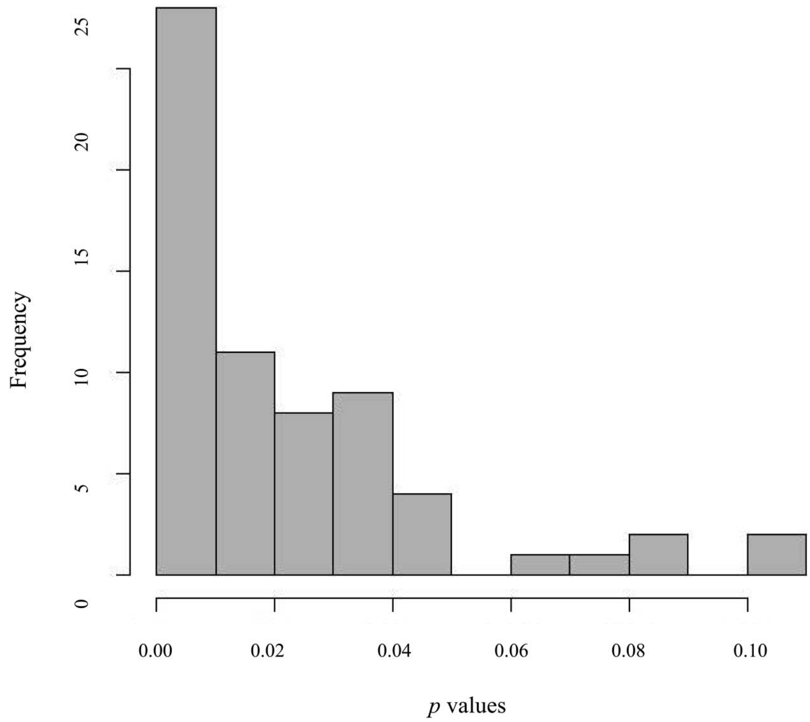


Figure 4. The p-distribution of studies from $p = 0$ to 0.11. The distribution of studies with p values from 0 to 0.11. In the case of unbiased reporting of studies testing a true effect, the distribution of p values is expected to be right-skewed and monotonically decreasing. This is what is observed in p-distribution of this meta-analysis ($p = .0009$).

measured in the Switch task (e.g., mutual exclusivity, Markman & Wachtel, 1988; fast mapping, Carey & Bartlett, 1978), it highlights that a key word-learning skill is in place by 12 months.

Similarly, performance did not differ as a function of the number of word-object pairings presented during habituation. That is, infants did not find it easier to learn one word-object association than to learn two or more in the same experiment, which presumably would make greater demands on working memory. This finding suggests that associative word learning is not merely influenced by the working memory demands present during the learning phase of the task. By 12 months of age, infants appear to be readily able to learn two words simultaneously, which reinforces our point above: infants may have strong word-object associative skills in place at the beginning of the second year of life.

Of interest to the authors, we found that bilingual infants, on average, perform better than their monolingual peers in the Switch task, consistent with some recent findings supporting a bilingual advantage in word learning tasks (e.g., Graf Estes & Hay, 2015; Singh, Fu, Tay, & Golinkoff, 2018). This advantage may be driven by the enriched phonological variability and multiple language inputs present in bilinguals' language environment (Byers-Heinlein & Fennell, 2014). Higher variability in speech input could help infants to find the most stable elements in their language environment as they develop their phonemic representations, which in turn facilitates the processes of word learning (Apfelbaum & McMurray, 2011). In addition, bilingual infants may possess better cognitive processing skills than monolinguals (Kovács & Mehler, 2009; Singh et al., 2015). Efficient information processing would help with the cognitive demands of word learning present in the Switch task (i.e., the need to simultaneously represent new words, new objects, and linking the two). Superior word learning skills could explain how bilinguals keep pace in overall vocabulary development with their monolingual peers, despite reduced input in each language (e.g., Core, Hoff, Rumiche, & Señor, 2013; but see Bilson, Yoshida, Tran, Woods, & Hills, 2015 for data inconsistent with this claim). However, we note that relatively few studies in our meta-analysis included bilingual infants ($n = 12$) and this result needs to be, therefore, confirmed with a larger sample. Nevertheless, current theoretical approaches to understanding early word learning do not necessarily predict a bilingual advantage in word learning (Curtin, Byers-Heinlein, & Werker, 2011); thus, our findings invite an updating of current theories.

Another important finding was that, in general, infants performed better when tested with dissimilar words than similar-sounding words (i.e., minimal pairs) in the Switch task. This confirms the findings of numerous articles that have focused specifically on this question (e.g., Pater, Stager, & Werker, 2004; Rost & McMurray, 2009; Stager & Werker, 1997; Thiessen, 2007). To successfully associate minimal pair words with objects, infants need to apply more refined and specific sensitivity to native language phonetic details than with dissimilar sounding words. As such, the demands of learning minimally different words are higher than that in dissimilar sounding words in the Switch task. However, we did not find an interaction between word similarity and age. Infants' ability to learn minimal pair words does not significantly change when they grow older, suggesting infants' phonological representations do not dramatically improve from 12 months to 20 months of age. This may indicate that infants' phonological representations are still somewhat fragile and not yet generalizable over most or all of the ages tested in our meta-analysis. Indeed, this is consistent with recent findings that even 17-month-old infants fail to learn similar-sounding words when slight deviations from typical pronunciations of the target phonemes are present (e.g., Fennell & Byers-Heinlein, 2014; Mattock et al., 2010). Further, recent studies have shown that it is not until 19 months that toddlers start to demonstrate more stable phonological representations, and can generalize the phonological representations across different realizations of words, such as recognizing familiar words across unfamiliar accents (Best, Tyler, Gooding, Orlando, & Quann, 2009). Our findings constrain theories of infant word learning that predict an eventual qualitative change in infants' phonological representations (Werker & Curtin, 2005), as we do not find evidence for such a change before age 20 months.

Despite the phonological representations perhaps not being completely solid at these ages, we did find evidence that infants' phonology is language-specific. Infants performed worse when tested on atypical word forms inconsistent with the sounds or sound patterns of the native language. This confirms previous reports that, by age 12 months, infants have developed clear expectations for what constitute reasonable object labels in their language (Mackenzie et al., 2012a, 2012b). Moreover, we found a near-significant trend that the effects of word-typicality change as infants grew older. Surprisingly, younger infants tended to show a stronger learning effect for typical over atypical words in comparison with older infants. This trend was unexpected because studies suggest that older infants exhibit a stronger selective bias for typical words (e.g., May & Werker, 2014; Vukatana, Curtin, & Graham, 2016). However, we note that there was a range of atypical words tested across studies, from slightly (e.g., accented words) to very (e.g., words with nonnative phonemes) atypical. An informal post hoc examination suggested that older infants (nine studies) performed worse than younger infants (14 studies) when tested with very atypical words, but older infants (15 studies) did better than younger infants (seven studies) when words were slightly atypical. Thus, slightly atypical words may be driving our age effect. This is consistent with Best et al. (2009)'s account of the emergence of phonological constancy around 19 months, which would allow older infants handle slight deviations from typical phonological structure but to reject very atypical words.

Finally, one of our goals was to investigate the role of different types of facilitative cues to adjudicate between two theoretical positions regarding the completeness of infants' early phonological representations. Low-level cues are proposed to enhance infants' initially incomplete phonological representations (e.g., Rost & McMurray, 2009). High-level conceptual cues are proposed to lighten the cognitive load of word learning, revealing infants' fairly well-established phonological representations (e.g., Fennell & Waxman, 2010). We found evidence that both types of cues boosted word learning, particularly for younger infants. We believe that this finding is consistent with the basic tenets of PRIMIR (Werker & Curtin, 2005), a language acquisition theory that focuses on the interaction between emergent phonology and the developing lexicon.

PRIMIR posits that infants encode and represent rich information from speech in terms of both phonetic details and phonologically irrelevant indexical cues (e.g., talker's gender). At the earliest word learning stage, young novice word learners are unsure which information is most important for word identity. Therefore, they tend to activate all information simultaneously, making the task of word association very demanding as both irrelevant and relevant information is highlighted. PRIMIR proposes three dynamic filters that direct infants' attention to key information in a word learning task: initial (or perceptual) biases, task demands, and developmental level. The task demands filter proposes that increased task difficulty reduces attention to relevant phonemic information in a word learning task. The high task demands of the relatively decontextualized Switch task substantially tax the cognitive capacity of very young infants. As a result, they are unable to direct their attention to the most relevant information in the task (e.g., phonemes). However, if we provide infants with facilitative cues (either low-level or high-level cues), we lessen the task demands of the Switch task, young infants have extra capacity to direct their attention to the phonemic details of the words to succeed in the Switch task. Both low-level cues and high-level cues allow infants to include more relevant information (i.e., phonological details) in their lexical representations, which subsequently facilitates infants' performance in the Switch task. Thus, our findings are in concordance with the propositions of PRIMIR.

However, our present article also points to an intriguing interaction between age and facilitative cues. Facilitative manipulations had a strong effect on younger infants' learning, but little or no effect on older infants' learning. One possibility is that older infants rely on different word learning cues than those typically present in the Switch task. According to the Emergentist Coalition Model (Golinkoff & Hirsh-Pasek, 2006), infants move from relying more on acoustic/perceptual cues when younger to an emphasis on social cues as they become experienced word learners. However, very few studies in our sample involved natural social-pragmatic cues (e.g., live speaker, eye gaze, and pointing). Thus, the facilitative cues provided in modified Switch tasks may not be sufficient to boost older infants' word-object associations. The interaction effect between age and facilitative cues may also be partially due to the greater number of studies using facilitative cues at younger ages ($n = 30$ at 12–15 months) than at older ages ($n = 13$ at 16–20 months). Further, we discovered that older infants were often tested on more difficult Switch task instantiations than younger infants. For example, 11 out 13 studies testing infants between 19 to 20 months used atypical words. Given that atypicality decreases performance as per our analysis, it is perhaps not surprising that older infants did not exhibit their potentially stronger associative learning. The use of challenging experimental conditions may be suppressing any boost given by facilitative cues.

On a more practical note, our moderator analyses reveal that researchers should take into account the stimuli, study design, and infants' language background when determining the target sample size of a Switch task study. Considering that the estimated power of significant studies in the current article averages to 48%, which is low but not atypical of infant research (Bergmann et al., 2018), it is especially important to consider the details of experimental designs to achieve an appropriately powered study. To facilitate this, we have shared the relevant data on the MetaLab website where researchers can determine estimated effect sizes based on different Switch experimental designs (e.g., select a larger sample size when testing learning of minimal pairs rather than dissimilar sounding words).

Questionable Research Practices (QRP)

We found evidence of a small study effect bias in the set of effect sizes, which can indicate publication and/or outcome reporting biases. We found no evidence for anomalous suppression of nonsignificant findings (i.e., the file-drawer problem); thus, this small study effect bias appears more attributable to QRPs that affect outcome reporting. We note that many Switch task articles were published before Simmons et al. (2011) brought an awareness of QRP to the forefront of the field.

One particular QRP concerned the use of one-tailed tests. Of course, one-tailed tests can be justified via a directional hypothesis in the Switch task: researchers expect infants who have habituated to exhibit a novelty preference (look longer to Switch over Same trial). However, the issue here was that only 2.8% of the studies in the meta-analysis used a one-tailed test and every one of those cases would not have reached statistical significance with a two-tailed test, suggesting a post hoc rather than a priori statistical decision. It is unsurprising that two-tailed tests are the norm, as it is possible for a Switch study to have a positive or negative significant effect. If infants find a habituation task too difficult, they can exhibit a familiarity rather than a novelty preference, as they are still processing the habituation stimulus at test (Hunter & Ames, 1988). Thus, the use of a one-tailed test implies that a researcher is confident that the task demands are appropriate for infants at that age, ruling out any possibility of a familiarity preference. This strong position seems unlikely in practice.

Another QRP that may have led to the small study effect is conducting analyses midway through experiments and then deciding to continue collecting data based on the results ("data peeking"). If an effect is marginally significant, researchers may increase the sample size to see if they achieve significance. This could result in larger samples having smaller effect sizes because studies with weak effect sizes are expanded (Schott et al., 2018). Indeed, a recent survey found that some infancy researchers report increasing their sample size as a function of obtained p values (Eason, Hamlin, & Sommerville, 2017). These issues underscore the challenges of working with difficult-to-recruit and difficult-to-test populations such as infants (average attrition rate in the Switch task = 29%) and the perceived publication pressures for statistical analyses to cross the $p < .05$ threshold, all of which can lead to QRP.

One concrete solution of QRP is to preregister research articles before data collection, an increasingly popular option in our field. Target sample sizes can be preregistered along with information such as confirmatory data analysis plans, using Websites such as the Open Science Framework (Lindsay, Simons, & Lilienfeld, 2016). Additional unregistered post hoc analyses can be conducted after data collection but would be clearly reported. Preregistration avoids the issue of reporting post hoc analyses as if they were planned and protects researchers from unwittingly changing analysis strategies over the course of exploring data. Further, transparent data peeking allows flexibility in sample size by applying appropriate statistical corrections (Schott et al., 2018). Another recommendation is expanding beyond the arbitrary, rigid, and binary cut-off criterion of $p < .05$ (Greenland et al., 2016). The p values do not provide key information (e.g., learning effect size, variability, and likelihood of replication). Changes in the field, including alternate statistical approaches (e.g., Bayesian Models: van de Schoot et al., 2014; use of confidence intervals: Cumming, 2014), can lessen the pressure to achieve and report statistically significant results. These exciting new tools and approaches, together with increasing QRP awareness, can improve the effect size biases seen in the field.

Limitations

Our meta-analysis was necessarily limited by the extant studies that have used the Switch task. This limits the degree to which we can generalize our findings beyond the types of questions that have been asked using the Switch task so far. For example, most minimal pair versions of the switch task involved stop consonants, with very few testing other types of phonemes. Thus, it is possible that some types of minimal pairs are easier to learn than others, a question that we were unable to investigate. Two of our moderators—low-level and high-level cues—encompass a range of manipulations, which would each be interesting to investigate individually. Relatedly, we coded most moderators in a binary way (e.g., typical vs. atypical), when such factors might better be thought of as operating on a continuum (e.g., the degree of typicality of a word). We also note that moderators were not randomly assigned to studies. Thus, it is possible that some relationships that we observed are confounded by unobserved variables. For example, we found that facilitative manipulations were of larger benefit to younger infants than older infants. However, it is also possible that researchers often tested older infants on a more difficult version of the Switch task, a point that we discussed earlier. This could mean that older infants and younger infants were not facing the same Switch task (the degree of difficulty was greater for older infants). While we did control for word typicality in our regression, it is still possible that there was an unmodeled interaction between typicality and facilitative manipulations (we did not have sufficient data for modeling this interaction), in that the demands of the task may exceed the facilitative effect of these manipulations and older infants, therefore, did not benefit.

Conclusion

The current study is the first to provide a meta-analysis of infant associative word learning in the lab. We examined data from 141 studies examining word-object associations via the Switch task conducted over the past 20 years. Rather than merely providing a review of current literature on the use of this infant habituation procedure, we used a quantitative meta-analytic approach to explore the efficacy of the task and its variants. The results of our study indicated a statistically significant and low to moderate effect size for an average infant in the word-object associative Switch task. Also, we examined the influences of some common, and oft-discussed, methodological variables on infants' Switch task performance. In general, we found no evidence that word learning in the task differed as a function of age, or how many word-object pairs were presented during habituation. On the other hand, we did find evidence that word learning is better for bilingual infants, for dissimilar-sounding words, and when facilitative manipulations are used. Finally, the current article showed evidence of bias in the reported effect sizes, which may be a result of questionable research practices. We note that awareness of these issues in the field has grown substantially since the first Switch task studies were conducted more than 20 years ago and have made several recommendations for continuing to improve research practices in this field.

References

Altvater-Mackensen, N., & Fikkert, P. (2010). The acquisition of the stop-fricative contrast in perception and production. *Lingua, 120,* 1898–1909. http://dx.doi.org/10.1016/j.lingua.2010.02.010

Apfelbaum, K. S., & McMurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science, 35,* 1105–1138. http://dx.doi.org/10.1111/j.1551-6709.2011.01181.x

Archer, S., Ference, J., & Curtin, S. (2014). Now you hear it: Fourteen-month-olds succeed at learning minimal pairs in stressed syllables. *Journal of Cognition and Development, 15,* 110–122. http://dx.doi.org/10.1080/15248372.2012.728544

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50,* 1088–1101. http://dx.doi.org/10.2307/2533446

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America, 109,* 3253–3258. http://dx.doi.org/10.1073/pnas.1113380109

Bergmann, C., & Cristia, A. (2016). Development of infants' segmentation of words from native speech: A meta-analytic approach. *Developmental Science, 19,* 901–917. http://dx.doi.org/10.1111/desc.12341

Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development, 89,* 1996–2009. http://dx.doi.org/10.1111/cdev.13079

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words. *Psychological Science, 20,* 539–542. http://dx.doi.org/10.1111/j.1467-9280.2009.02327.x

Bilson, S., Yoshida, H., Tran, C. D., Woods, E. A., & Hills, T. T. (2015). Semantic facilitation in bilingual first language acquisition. *Cognition, 140,* 122–134. http://dx.doi.org/10.1016/j.cognition.2015.03.013

Booth, A. E., Waxman, S. R., & Huang, Y. T. (2005). Conceptual information permeates word learning in infancy. *Developmental Psychology, 41,* 491–505. http://dx.doi.org/10.1037/0012-1649.41.3.491

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, West Sussex: Wiley. http://dx.doi.org/10.1002/9780470743386

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1,* 97–111. http://dx.doi.org/10.1002/jrsm.12

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science, 9,* 333–342. http://dx.doi.org/10.1177/1745691614529796

Brown, C., & Matthews, J. (1997). The role of feature geometry in the development of phonemic contrasts. In S. J. Hannahs & M. Young-Scholten (Eds.), *Focus on phonological acquisition* (pp. 67–112). Amsterdam: John Benjamins. http://dx.doi.org/10.1075/lald.16.05bro

Byers-Heinlein, K., & Fennell, C. T. (2014). Perceptual narrowing in the context of increased variation: Insights from bilingual infants. *Developmental Psychobiology, 56,* 274–291. http://dx.doi.org/10.1002/dev.21167

Byers-Heinlein, K., Fennell, C. T., & Werker, J. F. (2013). The development of associative word learning in monolingual and bilingual infants. *Bilingualism: Language and Cognition, 16,* 198–205. http://dx.doi.org/10.1017/S1366728912000417

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference, 15,* 17–29.

Chan, C. C., Tardif, T., Chen, J., Pulverman, R. B., Zhu, L., & Meng, X. (2011). English- and Chinese-learning infants map novel labels to objects and actions differently. *Developmental Psychology, 47,* 1459–1471. http://dx.doi.org/10.1037/a0024049

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Core, C., Hoff, E., Rumiche, R., & Señor, M. (2013). Total and conceptual vocabulary in Spanish-English bilinguals from 22 to 30 months: Implications for assessment. *Journal of Speech, Language, and Hearing Research, 56,* 1637–1649. http://dx.doi.org/10.1044/1092-4388(2013/11-0044)

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25,* 7–29. http://dx.doi.org/10.1177/0956797613504966

Curtin, S. (2009). Twelve-month-olds learn novel word-object pairings differing only in stress pattern. *Journal of Child Language, 36,* 1157–1165. http://dx.doi.org/10.1017/S0305000909009428

Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics, 39,* 492–504. http://dx.doi.org/10.1016/j.wocn.2010.12.002

Curtin, S., Campbell, J., & Hufnagle, D. (2012). Mapping novel labels to actions: How the rhythm of words guides infants' learning. *Journal of Experimental Child Psychology, 112,* 127–140. http://dx.doi.org/10.1016/j.jecp.2012.02.007

Curtin, S., Fennell, C., & Escudero, P. (2009). Weighting of vowel cues explains patterns of word-object associative learning. *Developmental Science, 12,* 725–731. http://dx.doi.org/10.1111/j.1467-7687.2009.00814.x

Dietrich, C., Swingley, D., & Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 16027–16031. http://dx.doi.org/10.1073/pnas.0705270104

Eason, A. E., Hamlin, J. K., & Sommerville, J. A. (2017). A survey of common practices in infancy research: Description of policies, consistency across and within labs, and suggestions for improvements. *Infancy, 22,* 470–491. http://dx.doi.org/10.1111/infa.12183

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629–634. http://dx.doi.org/10.1136/bmj.315.7109.629

Fais, L., Werker, J. F., Cass, B., Leibowich, J., Barbosa, A. V., & Vatikiotis-Bateson, E. (2012). Here's looking at you, baby: What gaze and movement reveal about minimal pair word-object association at 14 months. *Laboratory Phonology, 3,* 91–124. http://dx.doi.org/10.1515/lp-2012-0007

Fennell, C. T. (2006). Infants of 14 months use phonetic detail in novel words embedded in naming phrases. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development* (pp. 178–189). Boston, MA: Cascadilla Press.

Fennell, C. T. (2012). Habituation procedures. In E. Hoff (Ed.), *Research methods in child language: A practical guide* (pp. 3–16). Malden, MA: Wiley-Blackwell. http://dx.doi.org/10.1002/9781444344035.ch1

Fennell, C. T. (2012). Object familiarity enhances infants' use of phonetic detail in novel words. *Infancy, 17,* 339–353. http://dx.doi.org/10.1111/j.1532-7078.2011.00080.x

Fennell, C., & Byers-Heinlein, K. (2014). You sound like Mommy Bilingual and monolingual infants learn words best from speakers typical of their language environments. *International Journal of Behavioral Development, 38,* 309–316. http://dx.doi.org/10.1177/0165025414530631

Fennell, C. T., Byers-Heinlein, K., & Werker, J. F. (2007). Using speech sounds to guide word learning: The case of bilingual infants. *Child Development, 78,* 1510–1525. http://dx.doi.org/10.1111/j.1467-8624.2007.01080.x

Fennell, C. T., & Tsui, A. S. M. (2014). *Co-articulatory cues facilitate learning of minimal pairs*. Unpublished raw data.

Fennell, C. T., Tsui, A. S. M., & Hudon, T. M. (2016). Speech perception in simultaneously bilingual infants. In S. Montanari & E. Nicoladis (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 43–62). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/14939-004

Fennell, C. T., van der Feest, S., & Spring, M. (2012). *Consonant asymmetries at 14 months: Implications of acoustic variation on early phonological acquisition*. Unpublished manuscript, School of Psychology, University of Ottawa.

Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development, 81,* 1376–1383. http://dx.doi.org/10.1111/j.1467-8624.2010.01479.x

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech, 46*(2–3), 245–264. http://dx.doi.org/10.1177/00238309030460020901

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., . . . Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59,* 1–173. http://dx.doi.org/10.2307/1166093

Ference, J., & Curtin, S. (2015). The ability to map differentially stressed labels to objects predicts language development at 24 months in 12-month-olds at high-risk for Autism. *Infancy, 20,* 242–262. http://dx.doi.org/10.1111/infa.12074

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language, 51,* 419–439. Retrieved from https://www.jstor.org/stable/412864?seq=1#metadata_info_tab_contents

Fikkert, P. (2010). Developing representations and the emergence of phonology: Evidence from perception and production. In C. Fougeron, B. Kühnert, M. d'Imperio, & N. Vallée (Eds.), *Laboratory phonology 10: Variation, phonetic detail and phonological representation* (pp. 227–258). Berlin, Germany: Mouton.

Galle, M. E., Apfelbaum, K. S., & McMurray, B. (2015). The role of single talker acoustic variation in early word learning. *Language Learning and Development, 11,* 66–79. http://dx.doi.org/10.1080/15475441.2014.895249

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Golinkoff, R., & Hirsh-Pasek, K. (2006). Baby wordsmith: From associationist to social sophisticate. *Current Directions in Psychological Science, 15,* 30–33. http://dx.doi.org/10.1111/j.0963-7214.2006.00401.x

Graf Estes, K. (2012). Infants generalize representations of statistically segmented words. *Frontiers in Psychology, 3,* 447. http://dx.doi.org/10.3389/fpsyg.2012.00447

Graf Estes, K. (2014). Learning builds on learning: Infants' use of native language sound patterns to learn words. *Journal of Experimental Child Psychology, 126,* 313–327. http://dx.doi.org/10.1016/j.jecp.2014.05.006

Graf Estes, K., & Bowen, S. (2013). Learning about sounds contributes to learning about words: Effects of prosody and phonotactics on infant word learning. *Journal of Experimental Child Psychology, 114,* 405–417. http://dx.doi.org/10.1016/j.jecp.2012.10.002

Graf Estes, K., Edwards, J., & Saffran, J. R. (2011). Phonotactic constraints on infant word learning. *Infancy, 16,* 180–197. http://dx.doi.org/10.1111/j.1532-7078.2010.00046.x

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18,* 254–260. http://dx.doi.org/10.1111/j.1467-9280.2007.01885.x

Graf Estes, K. G., & Hay, J. F. (2015). Flexibility in bilingual infants' word learning. *Child Development, 86,* 1371–1385. http://dx.doi.org/10.1111/cdev.12392

Graf Estes, K. G., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy, 18,* 797–824. http://dx.doi.org/10.1111/infa.12006

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, *P* values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology, 31,* 337–350. http://dx.doi.org/10.1007/s10654-016-0149-3

Harrell, F. E. (2001). General aspects of fitting regression models. In F. E. Harrell (Eds.), *Regression modeling strategies* (pp. 11–40). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4757-3462-1_2

Hay, J. F., Graf Estes, K., Wang, T., & Saffran, J. R. (2015). From flexibility to constraint: The contrastive use of lexical tone in early word learning. *Child Development, 86,* 10–22. http://dx.doi.org/10.1111/cdev.12269

Hay, J. F., Pelucchi, B., Graf Estes, K., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology, 63,* 93–106. http://dx.doi.org/10.1016/j.cogpsych.2011.06.002

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327,* 557–560. http://dx.doi.org/10.1136/bmj.327.7414.557

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research, 1,* 491–507.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23,* 524–532. http://dx.doi.org/10.1177/0956797611430953

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32,* 402–420. http://dx.doi.org/10.1006/jmla.1993.1022

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33,* 630–645. http://dx.doi.org/10.1006/jmla.1994.1030

Kovács, Á. M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 6556–6560. http://dx.doi.org/10.1073/pnas.0811323106

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255,* 606–608. http://dx.doi.org/10.1126/science.1736364

Leon, A. C., & Heo, M. (2009). Sample sizes required to detect interactions between two binary fixed-effects in a mixed-effects linear regression model. *Computational Statistics & Data Analysis, 53,* 603–608. http://dx.doi.org/10.1016/j.csda.2008.06.010

Lindsay, S. D., Simons, D. J., & Lilienfeld, S. O. (2016). *Research preregistration 1010*. Retrieved from https://www.psychologicalscience.org/observer/research-preregistration-101#.WWkLDhPyuys

MacKenzie, H., Curtin, S., & Graham, S. A. (2012a). Class matters: 12-month-olds' word-object associations privilege content over function words. *Developmental Science, 15,* 753–761. http://dx.doi.org/10.1111/j.1467-7687.2012.01166.x

MacKenzie, H., Curtin, S., & Graham, S. A. (2012b). 12-month-olds' phonotactic knowledge guides their word-object mappings. *Child De-*

velopment, 83, 1129–1136. http://dx.doi.org/10.1111/j.1467-8624.2012.01764.x

MacKenzie, H., Graham, S. A., & Curtin, S. (2011). Twelve-month-olds privilege words over other linguistic sounds in an associative learning task. *Developmental Science, 14,* 249–255. http://dx.doi.org/10.1111/j.1467-7687.2010.00975.x

MacKenzie, H. K., Graham, S. A., Curtin, S., & Archer, S. L. (2014). The flexibility of 12-month-olds' preferences for phonologically appropriate object labels. *Developmental Psychology, 50,* 422–430. http://dx.doi.org/10.1037/a0033524

Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science, 18,* 387–391. http://dx.doi.org/10.1111/j.1467-9280.2007.01910.x

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20,* 121–157. http://dx.doi.org/10.1016/0010-0285(88)90017-5

Mattock, K., Polka, L., Rvachew, S., & Krehm, M. (2010). The first steps in word learning are easier when the shoes fit: Comparing monolingual and bilingual infants. *Developmental Science, 13,* 229–243. http://dx.doi.org/10.1111/j.1467-7687.2009.00891.x

May, L., & Werker, J. F. (2014). Can a click be a word? Infants' learning of non-native words. *Infancy, 19,* 281–300. http://dx.doi.org/10.1111/infa.12048

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science, 317,* 631. http://dx.doi.org/10.1126/science.1144073

MetaLab. (2016, June 26). *Interactive tools for community-augmented meta-analysis, power analysis, and experimental planning in language acquisition research*. Retrieved from http://metalab.stanford.edu/

Mitchell, C., & McMurray, B. (2009). On leveraged learning in lexical acquisition and its relationship to acceleration. *Cognitive Science, 33,* 1503–1523. http://dx.doi.org/10.1111/j.1551-6709.2009.01071.x

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., & the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151,* 264–269, W64. http://dx.doi.org/10.7326/0003-4819-151-4-200908180-00135

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105–125. http://dx.doi.org/10.1037/1082-989X.7.1.105

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349,* aac4716. http://dx.doi.org/10.1126/science.aac4716

Pater, J., Stager, C., & Werker, J. (2004). The perceptual acquisition of phonological contrasts. *Language, 80,* 384–402. http://dx.doi.org/10.1353/lan.2004.0141

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology, 61,* 991–996. http://dx.doi.org/10.1016/j.jclinepi.2007.11.010

Rabagliati, H., Ferguson, B., & Lew-Williams, C. (2019). The profile of abstract rule learning in infancy: Meta-analytic and experimental evidence. *Developmental Science, 22,* e12704. http://dx.doi.org/10.1111/desc.12704

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science, 12,* 339–349. http://dx.doi.org/10.1111/j.1467-7687.2008.00786.x

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy, 15,* 608–635. http://dx.doi.org/10.1111/j.1532-7078.2010.00033.x

Schott, E., Rhemtulla, M., & Byers-Heinlein, K. (2018, November). Should I test more babies? Solutions for transparent data peeking. *Infant Behavior and Development*. Advance online publication. http://dx.doi.org/10.1016/j.infbeh.2018.09.010

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R* (pp. 28–36). New York, NY: Springer. http://dx.doi.org/10.1007/978-3-319-21416-0

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359–1366. http://dx.doi.org/10.1177/0956797611417632

Singh, L., Fu, C. S., Rahman, A. A., Hameed, W. B., Sanmugam, S., Agarwal, P., . . . the GUSTO Research Team. (2015). Back to basics: A bilingual advantage in infant visual habituation. *Child Development, 86,* 294–302. http://dx.doi.org/10.1111/cdev.12271

Singh, L., Fu, C. S., Tay, Z. W., & Golinkoff, R. M. (2018). Novel word learning in bilingual and monolingual infants: Evidence for a bilingual advantage. *Child Development, 89,* e183–e198.

Sloutsky, V. M., & Robinson, C. W. (2008). The role of words and sounds in infants' visual processing: From overshadowing to attentional tuning. *Cognitive Science, 32,* 342–365. http://dx.doi.org/10.1080/03640210701863495

Smith, L. B., Jones, S. S., Yoshida, H., & Colunga, E. (2003). Whose DAM account? Attentional learning explains Booth and Waxman. *Cognition, 87,* 209–213. http://dx.doi.org/10.1016/s0010-0277(02)00236-6

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388,* 381–382. http://dx.doi.org/10.1038/41102

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76,* 147–166. http://dx.doi.org/10.1016/S0010-0277(00)00081-0

Thiessen, E. D. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language, 56,* 16–34. http://dx.doi.org/10.1016/j.jml.2006.07.002

Thiessen, E. D. (2011). When variability matters more than meaning: The effect of lexical forms on use of phonemic contrasts. *Developmental Psychology, 47,* 1448–1458. http://dx.doi.org/10.1037/a0024439

Thiessen, E. D., & Pavlik, P. I., Jr. (2016). Modeling the role of distributional information in children's use of phonemic contrasts. *Journal of Memory and Language, 88,* 117–132. http://dx.doi.org/10.1016/j.jml.2016.01.003

Thiessen, E. D., & Yee, M. N. (2010). Dogs, bogs, labs, and lads: What phonemic generalizations indicate about the nature of children's early word-form representations. *Child Development, 81,* 1287–1303. http://dx.doi.org/10.1111/j.1467-8624.2010.01468.x

Tincoff, R., & Jusczyk, P. W. (2012). Six-Month-Olds comprehend words that refer to parts of the body. *Infancy, 17,* 432–444. http://dx.doi.org/10.1111/j.1532-7078.2011.00084.x

Tsui, A. S. M., Fais, L., & Fennell, C. T. (2018). *Effects of referential cues on word learning*. Unpublished raw data.

Tsui, A. S. M., & Fennell, C. T. (2016). *Effects of cognate training on minimal pairs*. Unpublished raw data.

Tsuji, S., & Cristià, A. (2014). Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology, 56,* 179–191. http://dx.doi.org/10.1002/dev.21179

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development, 85,* 842–860. http://dx.doi.org/10.1111/cdev.12169

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36,* 1–48. http://dx.doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1,* 112–125. http://dx.doi.org/10.1002/jrsm.11

Vukatana, E., Curtin, S., & Graham, S. A. (2016). Infants' acceptance of phonotactically illegal word forms as object labels. *Journal of Child Language, 43,* 1400–1411. http://dx.doi.org/10.1017/S0305000915000707

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences, 13,* 258–263. http://dx.doi.org/10.1016/j.tics.2009.03.006

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72,* 655–684. http://dx.doi.org/10.1111/1467-8624.00304

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology, 34,* 1289–1309. http://dx.doi.org/10.1037/0012-1649.34.6.1289

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1,* 197–234. http://dx.doi.org/10.1080/15475441.2005.9684216

Werker, J. F., & Fennell, C. T. (2004). Listening to sounds versus listening to words: Early steps in word learning. In D. G. Hall & S. Waxman (Eds.), *Weaving a lexicon* (pp. 79–109). Cambridge: MIT Press.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy, 3,* 1–30. http://dx.doi.org/10.1207/S15327078IN0301_1

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7,* 49–63. http://dx.doi.org/10.1016/S0163-6383(84)80022-3

Zhao, Q. (2014). *How acoustic salience influences infants' word mapping* (Unpublished Master dissertation). University of Tennessee, Knoxville, TN.

Received March 21, 2018
Revision received December 4, 2018
Accepted December 18, 2018 ■