Bilingualism: Language and Cognition

cambridge.org/bil

Language and categorization in monolinguals and bilinguals

Sarah Fairchild[1] and Anna Papafragou[2]

[1]Department of Psychological & Brain Sciences, University of Delaware and [2]Department of Psychological & Brain Sciences, University of Delaware

Research Article

Cite this article: Fairchild S, Papafragou A (2020). Language and categorization in monolinguals and bilinguals. *Bilingualism: Language and Cognition* **23**, 618–630. https://doi.org/10.1017/S1366728919000488

Received: 29 April 2017
Revised: 6 June 2019
Accepted: 26 June 2019
First published online: 5 August 2019

Key words:
categorization; verbal labels; bilinguals; executive function

Address for correspondence:
Sarah Fairchild, E-mail: sarahcfairchild@gmail.com

Abstract

People assume that objects labelled alike belong to the same category. Here we asked whether the role of labels in categorization depends on individuals' language experience, linguistic abilities, and/or cognitive abilities. We compared monolinguals' and bilinguals' use of phonologically licit words (zeg), illicit words (gsz), and non-linguistic frames (in addition to a baseline condition with no additional cues) in forming novel categories. For both groups, licit words affected categorization more than frames, especially in the absence of perceptual evidence for category boundaries; illicit words also shifted categorization preferences compared to frames. Furthermore, linguistic abilities predicted reliance on both licit and illicit words, and bilingualism predicted reliance on illicit words in categorization. Thus, in both monolinguals and bilinguals, novel (and even unconventional) linguistic labels act as unique category markers but their use in categorization depends on individual language processing skills (and, in some cases, exposure to a second language).

Introduction

The ability to group entities into categories is one of the key cognitive abilities that enables humans to navigate an extremely complex world. Multiple dimensions of any given entity are integrated into the nuanced process of deciding how it should be categorized. It is widely recognized that language affects this process of categorization in concert with other perceptual and conceptual information. For instance, people are more likely to believe that two objects belong together when they share a label as compared to when no labels are present (Yamauchi & Markman, 2000; Sloutsky, Lo & Fisher, 2001; Lupyan, Rakison & McClelland, 2007; Johanson & Papafragou, 2016; Gervits, Johanson & Papafragou, 2016). In one study, Lupyan et al. (2007) showed that participants learned a labeled category of 'aliens' faster than an unlabeled category, regardless of whether the label was introduced visually or auditorily. Likewise, participants were more likely to group a novel entity with a like-labeled novel object than with a differently-labeled but equally perceptually similar object (Johanson & Papafragou, 2016). In this way, adults can use language as a tool to identify categories of novel entities (Gleitman & Papafragou, 2016; Wolff & Holmes, 2011).

The influence of language on category identification begins well before adulthood, with linguistic labels acting as 'invitations to form categories' for infants and children (Jones, Smith & Landau, 1991; Waxman & Markow, 1995; Balaban & Waxman, 1997; Waxman, 1999; Nazzi & Gopnik, 2001; Booth & Waxman, 2002; Oakes & Rakison, 2003; Gershkoff-Stowe & Smith, 2004; Welder & Graham, 2006; Fulkerson & Waxman, 2007; Plunkett, Hu & Cohen, 2008; Ferry, Hespos & Waxman, 2010; Fairchild, Mathis & Papafragou, 2018). For instance, 12–13-month-old infants identified the category of cars more effectively when the label "car" was introduced alongside a toy automobile compared to infants in a no-label control condition (Waxman & Markow, 1995). Similar results have been obtained with unfamiliar objects, with novel labels effectively encouraging 18-month-olds to extract new categories (Booth & Waxman, 2002). Furthermore, previous research suggests that language may play a unique role for categorization compared to other cues: for instance, words activate existing concepts more efficiently than associative cues do in adults (Boutonnet & Lupyan, 2015; Lupyan & Thompson-Schill, 2012) and have an advantage over tones (Ferry et al., 2010), content-filtered speech (Balaban & Waxman, 1997), and primate vocalizations (Ferry, Hespos & Waxman, 2013) in facilitating categorization in infants. However, it should be pointed out that, in previous work, linguistic and other cues (e.g., tones) have not always been equated in terms of salience or discriminability. Such a step would ensure clear support for the conclusion that language has a special role compared to various non-linguistic cues in promoting categorization.

Current literature leaves open the issue whether the effects of labels on categorization might depend on linguistic and cognitive characteristics of individuals. Some studies have noted in passing that adults' reliance on labels during categorization is variable (Sloutsky, Kloos &

Fisher, 2007; Sloutsky & Fisher, 2012), but have not explored the underlying reasons for this variability. A major open question is whether differences in an individual's linguistic experience – specifically, whether that experience involves one or multiple languages – might affect the way in which that individual uses linguistic labels to draw category boundaries. Monolinguals and bilinguals differ in many respects of their language use (e.g., Grosjean, 1989; Kroll, Bobb & Hoshino, 2014) but two specific aspects of the bilingual experience are particularly relevant for category formation.

First, bilingual children adhere less strictly to the one object-one name rule (known as the 'mutual exclusivity constraint'; Markman, 1991) than their monolingual peers, since bilingual children must learn to map labels in two languages onto their real-world referents (Davidson, Jergovic, Imami & Theodos, 1997; cf. also Davidson & Tell, 2005; Byers-Heinlein & Werker, 2009; Houston-Price, Caloghiris & Raviglione; 2010; but see Frank & Poulin-Dubois, 2002). Since categorization itself incorporates many different types of criteria, these results suggest that bilinguals might be more flexible than monolinguals in setting up connections between unfamiliar linguistic labels and the categorical structure of stimuli.

Second, and relatedly, experience with multiple languages has been associated with improved learning of form-meaning links for new words in adults (Bartolotti & Marian, 2012; Cenoz & Valencia, 1994; Sanz, 2000; Cenoz, 2003; Keshavarz & Astaneh, 2004; van Hell & Mahn, 1997). For example, Spanish–Basque bilinguals were shown to be at an advantage over Spanish monolinguals in acquiring English (Cenoz & Valencia, 1994). Van Hell and Mahn (1997) showed that experienced language learners outperformed novice language learners: in both the number of retained foreign words, and in the speed of their retrieval. Similarly, Papagno and Vallar (1995) found that bilinguals performed better on a foreign-word learning task than monolinguals. Another study found that bilinguals were better able to overcome competition in acquiring an artificial language than monolinguals (Bartolotti & Marian, 2012). Interestingly, bilinguals have a word learning advantage over monolinguals even when learning novel words with unfamiliar phonological properties (Kaushanskaya & Marian, 2009), or Morse code languages (Bartolotti, Marian, Schroeder & Shook, 2011). Furthermore, this advantage seems to be quite general, since it does not depend on the particular languages available to a bilingual person. In one study (Kaushanskaya & Marian, 2009), both early English–Spanish and early English–Mandarin bilinguals showed a word-learning advantage for phonologically unfamiliar words compared to English monolinguals, despite the fact that Spanish but not Mandarin is orthographically and phonologically close to English. If bilinguals' experience learning two linguistic systems makes them more flexible in learning novel linguistic stimuli, bilinguals may be influenced by unfamiliar words or word-like stimuli to a greater extent than monolinguals when forming new categories (the difference should not extend to monolinguals' and bilinguals' use of non-linguistic cues to form categories, and should not impact the overall advantage of language over non-linguistic cues in both groups). Current studies of the word learning advantage in bilingualism do not allow us to evaluate the role of language in category formation, because they have tested word learning mostly by pairing phonologically new word forms with translations to the meanings of known words in one of the bilingual participants' languages, as opposed to pairing them with referents from truly novel categories (e.g., Kaushanskaya & Marian, 2009).

Extending this line of reasoning, the use of novel words to affect categorization within the group of bilinguals might differ depending on a bilingual's specific language history. There is evidence that age of acquisition/length of exposure to a second language affects the word learning advantage in bilinguals: only early (but not late) English–Spanish bilinguals outperformed English monolinguals on a word learning task in an artificial language (Kaushanskaya & Marian, 2007). It is therefore possible that length of exposure and/or degree of proficiency in the second language might affect bilinguals' tendency or ability to rely on new words in drawing category boundaries. None of these factors should affect bilinguals' reliance on non-linguistic cues to form category boundaries.

Further linguistic factors might bear on the use of language to encourage category identification for both monolinguals and bilinguals. One possible source of variation is individual differences in verbal abilities such as language comprehension/production capabilities, vocabulary size, and verbal working memory (e.g., Daneman & Carpenter, 1980; King & Just, 1991; Tanner, Inoue & Osterhout, 2014; Tanner & Van Hell, 2014). We know that infants and children with larger vocabularies are more likely to categorize objects on the basis of a shared label than their less linguistically-skilled peers (Waxman & Markow, 1995; Nazzi & Gopnik, 2001; Ferguson, Havy & Waxman, 2015). It is likely that the relationship between verbal skills and use of language in categorization extends into adulthood.

An additional possible source of individual differences in the use of language in categorization is variation in cognitive abilities (e.g., Kane & Engle, 2002; Carlson, Moses & Claxton, 2004; Friedman, Miyake, Young, DeFries, Corley & Hewitt, 2008; Miyake & Friedman, 2012). When forming novel categories, adults must simultaneously take into account multiple dimensions of the entities at hand, including various perceptual, conceptual, and linguistic features, and inhibit irrelevant dimensions in order to make quick categorization decisions, drawing on our limited-capacity executive function (EF) system. In fact, categorization tasks are often used as measures of executive function (e.g., Bialystok, Barac, Blaye & Poulin-Dubois, 2010). Additionally, EF is recruited in learning words (de Abreu, Gathercole & Martin, 2011) that are then applied to objects and categories. Thus it seems reasonable to expect that individuals with better EF should be able to more efficiently use linguistic information to form new categories. EF is made up of several subcomponents (working memory, inhibitory control, and cognitive flexibility) that may individually support the relationship between language and categorization to varying degrees. However, currently, the precise role that cognitive abilities play in the use of language in categorization remains underspecified.

The role of executive function connects back to effects of bilingualism because it has been argued that, at least in children and older adults, bilinguals outperform monolinguals in a variety of non-linguistic executive function tasks (Bialystok, 1999; cf. Kroll et al., 2014; Valian, 2015, for reviews). More recently, failures to replicate have called the bilingual advantage into question (Paap & Greenberg, 2013). It thus remains an open issue whether potential differences in executive function can be identified between monolingual and bilingual groups, and whether any such differences themselves might impact the use of language in category learning.

The present study

In the present study, we address these open questions by investigating how English-speaking monolingual and early Spanish–

English and late English–Spanish bilingual adults use language and other cues to form new categories. The early Spanish–English bilinguals were primarily heritage language speakers who learned Spanish from birth but spoke English in educational and professional contexts from an early age. The late English–Spanish bilinguals were all native speakers of English who began learning Spanish in high school or college. The two groups allowed us to investigate differences due to specific types of bilingual language experience (discussed in more detail below). We focus on the domain of artifacts because this domain has been argued to be more susceptible to categorization by external cues (such as labels) than natural kinds, which tend to be categorized on the basis of 'essences' (e.g., Rhodes & Gelman, 2009; Diesendruck & Peretz, 2013; Gelman, 2013). We explore whether language has a UNIQUE role in shaping novel artifact categories in both monolingual and bilingual adults. Additionally, we ask whether the role of language and other cues in categorization is MALLEABLE due to bilingualism or other linguistic and cognitive characteristics that vary across individuals.

We adapted a classic triad task that has been used extensively in the developmental literature to study categorization (Gelman & Markman, 1986, 1987; Landau & Shipley, 2001; Gelman & Coley, 1990; Davidson & Gelman, 1990; Gelman & O'Reilly, 1988; Johanson & Papafragou, 2016; Fairchild et al., 2018; Sloutsky et al., 2001; Deng & Sloutsky, 2012; Gelman & Davidson, 2013). In this task, participants are presented with three novel objects and have to group one object (target) with one of the two remaining objects (standards). Critically, the target shares a label with one of the Standards, but the shared label does not always align with the degree of perceptual similarity between the target and each of the standards. This paradigm offers a strong test of the potential role of language for category formation since prior developmental work has shown that the degree to which children follow labels to draw category boundaries flexibly integrates perceptual considerations (e.g., Gelman & Davidson, 2013; Sloutsky et al., 2001).

In the present study we made three important modifications to this basic paradigm. First, we varied the type of cue used to group together the target and one of the standards. We compared (a) novel words that are phonologically licit in English and could serve as labels for the unfamiliar artifacts (e.g., *zeg*), (b) novel words that are phonologically illicit in English (e.g., *gxz*), (c) non-linguistic cues (patterned frames that surround each object), (d) and no cues. Second, we equated all cue types for discriminability, such that any advantage of one particular cue would be unlikely to be explained by lower-level visual differences among the cue types (e.g., two licit words being more visually distinct than two illicit words or frames). In this respect, we went beyond prior work with both children and adults that has compared linguistic to other cues in categorization, since this work has typically omitted this step. Third, we presented all cue types neutrally on the display next to the artifacts, instead of introducing the licit words (or any of the other cues) as "names" for the objects. Thus, any difference among the cues could not be attributed to differences in perceived intentionality (e.g., the labels appearing to be the intended names for the objects as given by the experimenter), and we are able to measure what types of cues participants spontaneously make use of.

We reasoned that if labels have a unique ability to encourage category formation, both monolingual and bilingual participants would be more likely to group together objects when they share a licit word compared to an illicit word, frame, or no cue at all. We predicted that the effects of licit words would be particularly striking when the exemplars sharing a licit word are sufficiently perceptually dissimilar to each other and thus would not otherwise be grouped together (cf. Gelman & Davidson, 2013; Sloutsky et al., 2001). In addition, we anticipated that bilingual adults might be more likely than monolingual adults to rely on licit words to categorize since bilinguals learn novel words more easily than monolinguals (and might be more flexible in adopting newly-introduced connections between words and categories). This prediction might further extend to illicit words, which can be considered in-between cases that are not linguistic in the sense that they could be words, but are 'language-like' as they still bear the visual features of words. Since the bilingual word learning advantage extends to words that are phonologically ill-formed in the languages spoken by bilinguals (Kaushanskaya & Marian, 2009), we reasoned that bilingual participants might be more likely to accept phonologically unfamiliar strings as novel names for unfamiliar objects, but that the use of the non-linguistic frames would be unaffected by language experience.

Recall that bilinguals seem to have a general advantage in learning new words compared to monolinguals regardless of their specific languages; in fact, this advantage has been found when bilinguals speaking closely related languages such as English and Spanish were compared to English monolinguals (Kaushanskaya & Marian, 2009). Here the choice of two different bilingual groups allowed us to test whether reliance on novel linguistic cues during categorization varied with length of exposure to a second language (English for the early bilinguals and Spanish for the late bilinguals). One possible hypothesis we considered was that early exposure ensures an extensive period of time when both languages are used and activated in parallel and would therefore increase the likelihood that words become a flexible tool for categorizing unfamiliar exemplars: thus, only early (but not late) bilinguals would show differential sensitivity to licit (and perhaps also illicit) words in categorization compared to English monolinguals. According to a second possible hypothesis, successful acquisition of a second language even later in life might affect the ease of acquiring novel labels and applying them to unfamiliar categories, especially when the second language is acquired to a highly proficient level. If so, reliance on linguistic cues during category formation would be higher in both bilingual groups in our sample, regardless of length of exposure, compared to English monolinguals.

To test for the potential relationship between verbal abilities and use of language in category formation, participants completed a lexical decision task. This task has been shown to accurately predict language processing skills in various groups of monolinguals and bilinguals (Harrington, 2006). Therefore, we predicted that individuals with higher scores on the lexical decision task would be better able to benefit from licit words (but not non-linguistic cues) in categorization (cf. also Waxman & Markow, 1995; Nazzi & Gopnik, 2001; Ferguson et al., 2015). Another question of interest was whether this relationship exists within the monolingual group, given that some previous work has reported a disadvantage for bilinguals as compared to monolinguals on the lexical decision task (e.g., Soares & Grosjean, 1984).

Participants also completed several widely used cognitive measures that include aspects of executive function: (i) the digit span task targeting verbal working memory (Diamond, 2013); (ii) the Corsi block-tapping task targeting visual-spatial working memory (Kessels, Van Zandvoort, Postma, Kappelle & De Haan, 2000; Fischer, 2001), versions of which play an important role in clinical diagnoses of diseases such as Alzheimer's (Carlesimo et al., 1994),

Table 1. Means and standard deviations (in italics) of participant characteristics.

Participant Type	Gender	Age	Self-Rated English Proficiency	Self-Rated Spanish Proficiency
English Monolingual	13 Male 19 Female	18.53 *.76*	9.49 *.60*	—
English–Spanish	3 Male 13 Female	18.94 *1*	9.84 *.35*	6.84 *2.08*
Spanish–English	6 Male 10 Female	19.13 *.96*	9.42 *1.1*	8.64 *.88*
Total	9 Male 23 Female	19.03 *.97*	9.63 *.83*	7.74 *1.82*

Parkinson's (Stoffers, Berendse, Deijen & Wolters, 2003), and schizophrenia (Chey, Lee, Kim, Kwon & Shin, 2002); (iii) the Flanker test targeting inhibitory control (Eriksen & Eriksen, 1974) that is characterized by high test-retest reliability (Wöstmann, Aichert, Costa, Rubia, Möller & Ettinger, 2013); and (iv) the Wisconsin Card-Sorting test targeting cognitive flexibility that has been widely used to probe shifts between mental sets in children (Bull & Scerif, 2001; Romine, Lee, Wolfe, Homack, George & Riccio, 2004), adults (Rhodes, 2004; Buchsbaum, Greer, Chang & Berman, 2005), and patient populations (Gold, Carpenter, Randolph, Goldberg & Weinberger, 1997; Stuss, Levine, Alexander, Hong, Palumbo, Hamer, Murphy & Izukawa, 2000). Results from these cognitive measures were then related to the use of linguistic and non-linguistic cues in the triad task. Of interest was whether specific cognitive abilities are related to the ability to integrate different types of information in categorization. For example, because the cues disappeared before the response screen in the present experiment, we reasoned that verbal working memory might most strongly predict label use in categorization as participants may need to hold the licit or illicit words in working memory before making a response. Visuo-spatial working memory was predicted to have no such effect on licit or illicit word use in categorization but to potentially be predictive of frame use, since frames are visual cues that also need to be stored before a response can be made. We predicted that inhibitory control might be associated with performance on the categorization task, because making use of any of the cues on perceptually dissimilar trials requires inhibition of salient visual information. Similarly, to the extent that cognitive flexibility requires inhibition, we reasoned that it would be possible for cognitive flexibility to predict overall cue use. Neither inhibitory control nor cognitive flexibility should be particularly tied to the use of licit word labels over frames, however. Finally, we were interested to explore the extent to which differences in executive function exist between monolinguals and bilinguals, with corresponding effects on category learning.

Method

Participants

Thirty-two English monolingual adults and thirty-two English-dominant bilingual adults aged 18–21 participated. All participants were undergraduate students at the University of Delaware, and received course credit for their participation. In the bilingual group, there were sixteen Spanish–English bilinguals who learned Spanish as a native language but spoke English in educational and professional contexts (i.e., heritage language speakers of Spanish). The other sixteen participants in that group were English–Spanish bilinguals, who learned English as a native language and Spanish as a second language (L2) later in life (beginning in high school or college, i.e., adult L2 learners of Spanish). All bilinguals reported English as being their dominant language.

A language history questionnaire was administered in the beginning of the testing session and assessed self-rated English (and Spanish, for bilingual participants) proficiency (from 1–10, 10 being the highest). Groups did not differ in terms of self-rated English proficiency. As might be expected, within the bilingual group, Spanish–English bilinguals rated themselves as having significantly higher Spanish proficiency than the English–Spanish bilinguals, $t(30) = 3.175$, $p = .003$. Participant characteristics are presented in Table 1.

Procedure

Participants were tested individually or in pairs in a quiet room after they completed the language history questionnaire. All tasks were administered on a laptop with a 15.4" screen, approximately 24 inches away from the participant. The session lasted approximately one hour. All participants were tested in English. Each task was administered in the order below, with the exception of the linguistic measure that was administered last. The categorization and digit span tasks were presented using the OpenSesame experimental presentation software (Mathôt, Schreij & Theeuwes, 2012). The rest of the tasks were drawn from the PEBL experimental software test battery (Mueller & Piper, 2014).

Categorization task

Thirty-two grayscale photographs of objects chosen to be unfamiliar to participants were used as stimuli. These objects included arcane tools and pieces of other man-made devices. To ensure novelty, participants were asked at the completion of the experiment if they recognized any of the objects. The vast majority of participants could not name any of the stimuli, and responded that the objects appeared to be tools of some kind. Only one participant correctly identified a strawberry huller.

The 32 novel objects were divided into 16 pairs, with one object in the pair designated as standard A, and the other as standard B. Each pair was morphed together using the Fantamorph program. For each pair, five morphed pictures were created at 10%, 30%, 50%, 70%, and 90% similarity to standard A according to the program (see Figure 1 for an example).

Eight novel English-like licit words were generated using the ARC Nonword Database (Rastle, Harrington & Coltheart, 2002) with the following specifications: a length of three to five letters, orthographically existing onsets, orthographically existing bodies,

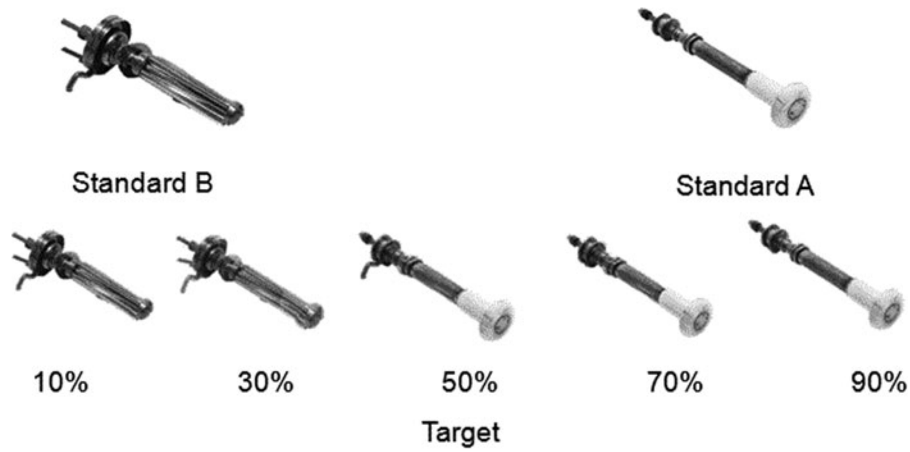

Fig. 1. Example stimuli. Percentages reflect target similarity to Standard A.

and legal bigrams (e.g., *zeg*, *wob*, *sten*). All labels were also created such that they were also phonologically possible labels in Spanish. Eight novel English-like illicit words were also created by taking the reverse of the licit words and replacing the vowels with x's. For example, *zeg* would become *gxz*. Replacing vowels with consonants ensured that these illicit words were phonologically disallowed in both English and Spanish. Since bilinguals are able to recognize phonotactic constraints in both of their languages (Sebastián-Gallés & Bosch, 2002), they should have no issues distinguishing licit from illicit words. Indeed, neuroimaging work indicates that strings of consonants are processed differently from both real words and phonologically possible novel words (Petersen, Fox, Snyder & Raichle, 1990; Dehaene, 1995). Finally, eight patterned frames (e.g., solid, dashed) were created to surround the objects. These served as non-linguistic cues. Examples of all cue types can be seen in Figure 2.

Two norming studies were conducted. First, we assessed whether or not people were sensitive to the similarity manipulations in our morphed stimuli. A separate group of 10 monolingual English-speaking adults were presented with all possible standard/target triads, and were asked to rate the similarity of the target to one of the standards on a 9-point scale, 10–90%. Participants were very accurate at rating the 30%, 50%, and 70% morphs; ratings did not differ significantly from actual similarity (all p's > .05). Ratings for the 10% and 90% stimuli differed significantly from actual similarity, but in the predicted direction. The 10% objects were rated significantly lower than the 30% objects, $t(9) = 7.003$, $p < .001$, and the 90% objects were rated significantly higher than the 70% objects. $t(9) = 10.902$, $p < .001$. Second, we ensured that all three cue types were equally discriminable. A separate group of 10 monolingual adults from the same population were asked to rate the discriminability of pairs of licit words, illicit words, and frames. For each pair, they answered the question 'How similar are these two items?' on a seven-point scale. Ratings did not differ across cue types (p's > .1).

For each trial of the categorization task, we combined standard A and B from one of the 16 novel object pairs, and a target from that set. The standards were presented on the top of the screen their placement (left side or right side) counterbalanced, and the target was centered on the bottom of the screen. The triad remained on the screen for 1000 milliseconds, at which point cues were introduced. There were four within-subjects cue conditions: licit words, illicit words, frames, and a no cue control condition. Cues appeared below the objects for 2000 milliseconds (see Figure 2), and the cue for the target always matched the cue for Standard A. When the cues disappeared, a red frame appeared along the edges of the screen.

At the beginning of the task, participants were given on-screen instructions that were read aloud by the experimenter. They were told that they would see groups of three objects, and that they would be given time to look at the objects. Then, a red frame would appear on the screen. They were told that, when the red frame appeared, they should press a key to indicate where they thought the bottom object belonged, either with the object on the right or the object on the left. Participants were instructed to make a response as quickly as possible.

There were four blocks of trials, one for each cue type, with block order counterbalanced across participants. Stimuli were fully rotated across cue types. Within each block, participants saw all possible target/standard triads in a random order, 20 trials per block, for a total of 80 trials.

Linguistic measure: Lexical decision task

Participants completed a lexical decision task targeting English comprehension. Twenty English words (e.g., EDUCATION) and twenty pseudowords (e.g., EMUCATION) were presented in a random order for 500 milliseconds each. Participants were asked to indicate by keypress, as quickly as possible, whether or not each stimulus was a real English word. Both accuracy and reaction times were collected. For the analysis of reaction times, incorrect responses and responses slower than 3000 milliseconds were excluded.

Cognitive measures: Digit span task

Participants completed the digit span task as a measure of verbal working memory. In a classic forward digit span task, participants view streams of numbers presented one at a time on the screen (e.g., "1, 8, 3, 4") and must later recall the numbers in the order they were given. We implemented a variant in which the numbers must be recalled in numerical order from smallest to largest (e.g., "1, 3, 4, 8"). The advantage of this type of digit span task over a forward or backward digit span is that participants are not only required to hold items in memory, but are also required to actively perform an operation on the number set (re-ordering). Because of this requirement, the re-ordered digit span is thought to more accurately represent verbal working memory than forward or backward digit span tasks that can be completed using short-term memory alone, not WORKING memory (Diamond,

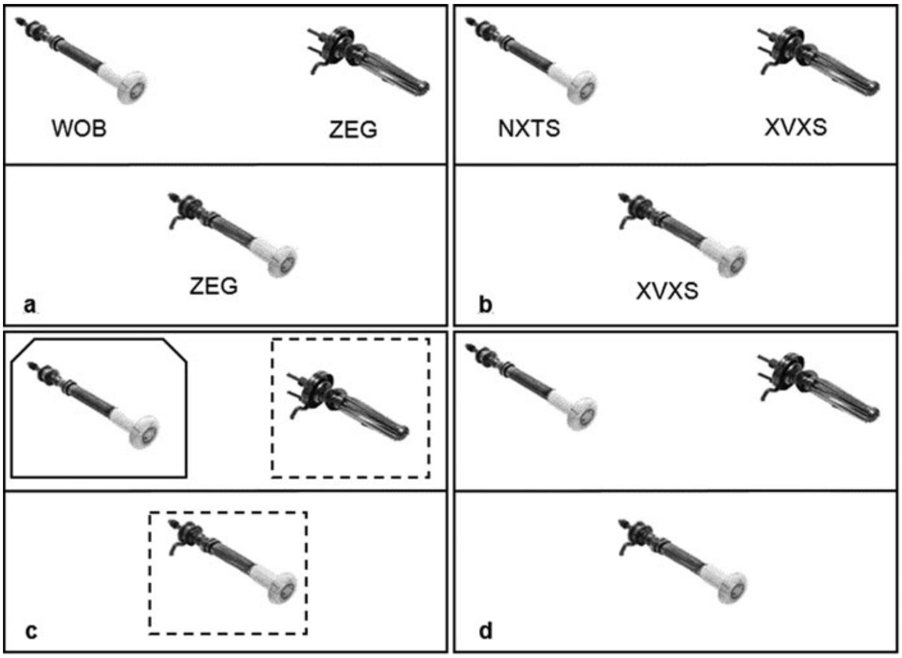

Fig. 2. Example stimuli for (a) Licit Word, (b) Illicit Word, (c) Frame, and (d) No Cue blocks.

2013). On each trial of the re-ordered digit span task in the present study, two to eight single-syllable digits (1, 2, 3, 4, 5, 6, 8, 9) were presented for 750 milliseconds each. Participants were then asked to type as many numbers as they could remember, in numerical order from smallest to largest (e.g., "1, 8, 3, 4" becomes "1, 3, 4, 8"). The recall phase was untimed, and participants could move on to the next trial when ready. All participants completed a total of 10 trials.

Cognitive measures: Corsi block-tapping task

The Corsi block-tapping task was given as a measure of visual-spatial working memory. On each trial of the task, an array of blue squares was presented on the screen, and then three to eight squares briefly sequentially changed color for 250 milliseconds each. Participants were asked to repeat what they had seen by clicking on the squares that changed color, in the order that they did so. There was no time constraint for responses. The task increased in difficulty on each trial and ended when the participant incorrectly answered three trials in a row. A measure of working memory span, the highest number squares accurately recalled (Cowan's k), was calculated for each participant.

Cognitive measures: Flanker task

In this inhibitory control task, participants responded to the direction of an arrow in the center of the screen, either left or right, by pressing a key as quickly as possible. The critical manipulation is the direction of other arrows surrounding the center arrow. On incongruent trials, the arrow was flanked by arrows facing the opposite direction (<<><<). On congruent trials, all of the arrows faced the same direction (<<<<<). On neutral trials, the center arrow was flanked by dashes (--<--). Single trials were also included, in which the arrow was presented in isolation (>). Inhibitory control is recruited to a greater extent for incongruent trials, which require participants to ignore the direction of distractor stimuli and respond only to the center arrow. Participants completed a total of 48 trials (12 per condition), preceded by 16 practice trials that included feedback (correct/incorrect). Trials timed out after 800 milliseconds if no response was detected. A measure of Flanker Conflict was calculated as the difference in reaction times between incongruent and congruent trials. Thus, participants with less Flanker Conflict had better inhibitory control.

Cognitive measures: Wisconsin Card Sorting task

A variant of the 64-card Wisconsin Card Sorting task (Kongs, Thompson, Iverson & Heaton, 2000) was administered as a measure of cognitive flexibility. Participants were given a deck of 64 cards, one card at a time, and were asked to sort the cards into four piles by clicking on the intended pile. There was no time constraint on responses. The cards could be sorted by color, number, or shape. Participants were not told the rule, but were given feedback (correct/incorrect) after each card they placed, allowing them to learn the rule. The sorting rule changed throughout the task, and participants were informed that if this happened (i.e., they started getting negative feedback), they should attempt to figure out the new rule as soon as possible. The rule changed up to four times over the course of the task. We calculated a measure of cognitive flexibility by taking the average number of trials it took participants to learn each new rule.

Results

Categorization in monolinguals and bilinguals

In a first analysis, we tested for potential differences in how monolinguals and bilinguals used different types of cues in categorization. We performed a 4 (Cue type: licit word, illicit word, frame, no cue) by 5 (Perceptual similarity: 10%, 30%, 50%, 70%, 90%) by 2 (Group: monolingual, bilingual) mixed ANOVA with cue compliance as the dependent variable. We considered cue-compliant responses to be those where the participant categorized the target object with the cue-matched Standard A. Cue compliance in the no cue block was calculated in the same way (i.e., in terms of responses to the predetermined standard A).

This analysis revealed a significant main effect of cue type, $F(3, 186) = 20.237$, $p < .001$, $\eta^2 = .068$. Bonferroni-corrected post-hoc analyses showed higher cue compliance for licit words ($M = .68$, $SD = .14$) as compared to illicit Words ($M = .61$, $SD = .13$), frames ($M = .57$, $SD = .12$), and the no cue condition ($M = .54$, $SD = .05$), all p's < .001. Cue compliance was also higher for the illicit word condition as compared to the frame, $p = .027$, and no cue conditions, $p < .001$. The difference between the frame and no cue conditions was not significant, $p > .1$. There was also a significant main effect of perceptual similarity, $F(4, 248) = 1068.208$, $p < .001$, $\eta^2 = .782$, such that cue compliance increased as the visual similarity of the objects increased. There was no significant main effect of group, $F(1, 62) = 3.120$, $p = .082$, $\eta^2 = .006$: cue compliance was similar for bilinguals ($M = .62$, $SD = .14$) and monolinguals ($M = .59$, $SD = .11$).

Finally, the analysis revealed a significant interaction between cue type and perceptual similarity, $F(12, 744) = 7.478$, $p < .001$, $\eta^2 = .051$. On the most dissimilar (10% similarity) trials, there was greater cue compliance for licit words as compared to frames, $p = .001$, and the no cue condition, $p < .001$; illicit words did not differ significantly from licit words, frames, or the no cue condition, all p's > .1. The same pattern was observed for 30% trials, with higher cue compliance observed for licit words as compared to frames and the no cue condition, both p's < .001 – but not the illicit words condition, $p > .1$. On trials that were ambiguous with respect to perceptual similarity (50% similarity), cue compliance was higher for licit words as compared to the no cue condition, $p < .001$; no other comparisons were significant. Cue compliance did not differ across cue types for the 70% and 90% similar trials. On these trials, cue compliance was at ceiling and therefore the cues did not alter responses above and beyond perceptual similarity. No other interactions reached significance (all p's > .1). In sum, then, regardless of participants' language background, licit words affected categorization compared to the no cue control and the frame condition when perceptual similarity between exemplars was low (10% and 30% similarity trials) or ambiguous (50% trials). The results from the monolingual and bilingual group are depicted in Figures 3 and 4.

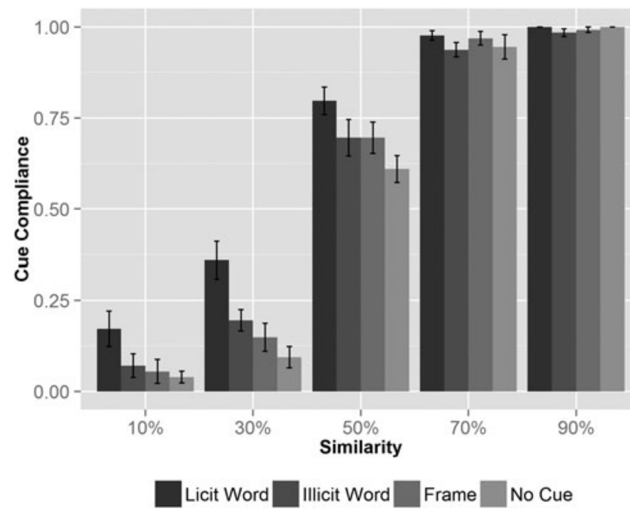

Fig. 3. Cue compliance for monolinguals in the Categorization task. Error bars represent +/− 1 S.E.M.

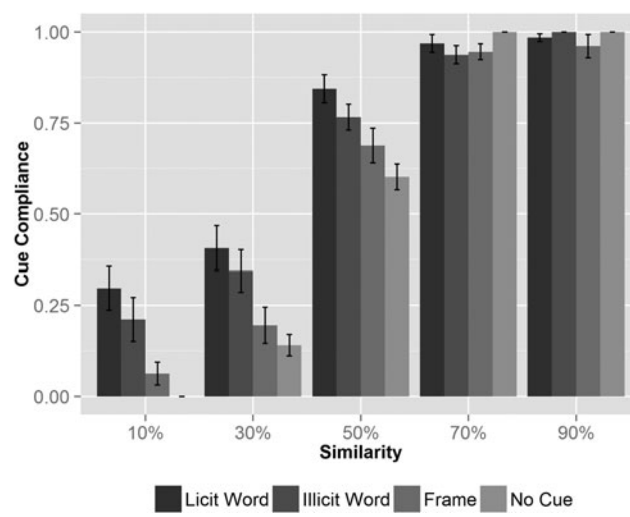

Fig. 4. Cue compliance for bilinguals in the Categorization task. Error bars represent +/− 1 S.E.M.

Types of bilinguals and categorization performance

In a second analysis, we targeted the role of specific types of bilingual language experience on the use of various cues in categorization. Focusing solely on data from the bilingual group, we performed a 4 (Cue type: licit word, illicit word, frame, no cue) by 5 (Perceptual similarity: 10%, 30%, 50%, 70%, 90%) mixed ANOVA with bilingual type (Spanish–English, English–Spanish) as a between-subjects factor using the proportion of cue-compliant responses made by the bilingual participants as the dependent variable.

Unsurprisingly, the analysis did reveal a significant main effect of cue type, $F(3, 90) = 11.783$, $p < .001$, $\eta^2 = .081$. There was greater cue compliance in the licit word condition ($M = .70$, $SD = .38$) as compared to the frame ($M = .57$, $SD = .43$), $p < .001$, and no cue ($M = .55$, $SD = .44$) conditions, $p < .001$ (the frame and no cue conditions did not differ, $p > .1$.) The licit word condition did not differ from the illicit word ($M = .65$, $SD = .40$) condition, $p > .1$; furthermore, there was greater cue compliance in the illicit word condition as compared to the frame, $p = .003$, and no cue conditions, $p < .001$.

The analysis also showed a significant main effect of perceptual similarity, $F(2.4, 72) = 409.207$, $p < .001$, $\eta^2 = .743$.[1] Cue compliance increased with each level of perceptual similarity (all p's < .001) except for the 70% and 90% conditions, $p > .1$. The analysis found no significant main effect of bilingual type, and no significant interactions between bilingual type and cue type or bilingual type and similarity (all p's > .1).

Finally, the analysis revealed a significant interaction between cue type and perceptual similarity, $F(6.6, 198) = 5.346$, $p < .001$, $\eta^2 = .069$. Post-hoc tests indicated that, for 10% similar trials, there was higher cue compliance in the licit word condition as compared to the frame, $p < .001$, and no cue conditions, $p < .001$; cue compliance was also higher for the illicit word than for the frame, $p = .020$, and the no cue conditions, $p = .002$ (no other differences were significant). On 30% similar trials, there was significantly greater cue compliance in the licit word

[1]Mauchly's test indicated that the assumption of sphericity had been violated for perceptual similarity, $\chi^2(9) = 37.831$, $p < .001$, and the interaction between perceptual similarity and cue type, $\chi^2(77) = 149.500$, $p < .001$, thus Greenhouse-Geisser corrections were applied to the degrees of freedom ($\varepsilon = .600$, and $\varepsilon = .550$, respectively).

Table 2. Means and standard deviations for individual differences measures. *$p < .05$ **$p < .01$, comparison between groups.

Variable	Monolinguals	Bilinguals	Total
Linguistic measures			
English Lexical Decision (ACC)	.90(*.09*)	.92(*.08*)	.91(*.08*)
English Lexical Decision (RTs)	673.88(*135.71*)	665.88(*129.20*)	665.17(*131.83*)
Cognitive measures			
Verbal Working Memory (%)	.77(*.22*)	.73(*.20*)	.75(*.21*)
Visual-Spatial Working Memory Span	5.75(*.83*)	5.67(*.90*)	5.71(*.86*)
Inhibitory Control (Conflict)	42.36(*42.14*)	58.72(*39.81*)	50.26(*41.52*)
Cognitive Flexibility (# trials)*	10.21(*3.80*)	12.99(*7.00*)	11.60(*5.71*)

condition as compared to the frame, $p = .006$, and no cue conditions, $p < .001$, which did not differ from one another, $p > .1$. There was also greater cue compliance for the illicit word as compared to the frame, $p = .020$, and no cue conditions, $p = .008$. Licit words and illicit words did not differ from one another, $p > .1$. For 50% similar (ambiguous) trials, there was again greater cue compliance for licit words than for frames, $p = .008$, and the no cue control, $p < .001$; the last two conditions did not differ, $p > .1$. Illicit word cue compliance was greater than the no cue condition, $p = .003$, but did not differ significantly from licit words or frames, both p's $> .1$. Cue compliance did not vary across conditions for 70% and 90% similar trials. In sum, within the bilingual group, we observed an advantage for both licit and illicit words over frames and the no cue control in the 10% and 30% similarity trials (licit and illicit word cue compliance did not differ from one another in those two cases). For the perceptually ambiguous (50%) trials, licit and illicit words had greater cue compliance than the no cue condition. Cue compliance was already at ceiling in the 70% and 90% conditions.

Individual variation in linguistic and cognitive measures

One of our primary goals in the present paper was to investigate how individual variation in cognitive and linguistic abilities – regardless of monolingual/bilingual status – affects the relationship between language and categorization. We used multiple regression analyses to address this question. As can be seen in Table 2, the two groups were similar, except that bilinguals performed significantly worse on the cognitive flexibility task, $t(62) = 2.071$, $p = .043$. Due to this difference, we treated cognitive flexibility as a control variable in the following analyses.

For these analyses, we calculated measures of CUE IMPACT, which represented the effectiveness of a particular cue in comparison to the no cue baseline. For each cue, we collapsed across levels of perceptual similarity and subtracted cue compliance in the no cue condition from cue compliance in each cue condition to obtain the cue impact (e.g., licit word compliance – no cue compliance = licit word impact). We then performed three separate multiple regression analyses, with the licit Word Impact, the illicit word impact, and the frame impact as dependent variables. Data from all participants were included (with the exception of one bilingual who did not complete the battery of individual differences measures). English Lexical Decision performance (accuracy and reaction times), Verbal Working Memory, Visual-Spatial Working Memory, Inhibitory Control, and Cognitive Flexibility were included in the model as predictors. Group (monolingual, bilingual) was also included as a factor to potentially uncover further differences between the groups.

This model significantly predicted licit Word Impact, $F(7, 54) = 2.592$, $p = .022$, $R^2 = .155$, and illicit word impact, $F(7, 54) = 3.127$, $p = .007$, $R^2 = .192$, but not frame impact ($p > .1$). Full results of each analysis are presented in Table 3. Closer inspection reveals that English Lexical Decision performance – both in terms of accuracy ($\beta = .411$. $t(54) = 3.085$, $p = .0321$) and reaction times ($\beta = -.272$. $t(54) = 2.027$, $p = .048$) – was a significant predictor of the licit word impact. That is, the faster and more accurately participants responded in the Lexical Decision task, the more likely they were to use licit words to form categories over and above what they would do in the no cue condition. Additionally, English Lexical Decision task accuracy was a significant predictor of illicit word impact ($\beta = .295$. $t(54) = 2.275$, $p = .027$), suggesting that better linguistic ability increased the use of any "language-like" cues in categorization. Interestingly, bilingualism status was also a significant predictor of illicit word Impact ($\beta = .333$. $t(54) = 2.710$, $p = .009$), such that bilinguals were more likely to be influenced by the illicit words than monolinguals. This difference did not emerge in the omnibus ANOVA comparing the two groups. Notice, however, that when the bilingual data were analyzed separately for purposes of comparing types of bilinguals, both licit and illicit words were found to shape categorization and to the same degree. Implications of these results are discussed in the next section.

Discussion

A wealth of prior research has demonstrated that infants, children, and adults tend to group novel objects together if these objects share the same linguistic label (e.g., Gelman & Markman, 1986; Waxman & Markow, 1995; Lupyan et al., 2007; Fairchild et al., 2018). This work has left open the question whether experience with more than one language might impact the process of using language to inform boundaries for novel categories. More generally, previous work has noted some variation in how individuals make use of language during categorization of novel exemplars (e.g., Sloutsky & Fisher, 2012) but has not systematically examined the nature of this variation.

Here we addressed these gaps in the literature by comparing the effects of novel (phonologically) licit words (*zeg*), illicit words (*gsz*) and non-linguistic cues (frames) on the way monolingual (English-speaking) and bilingual (early Spanish–English and late English–Spanish) adults identify unfamiliar artifact categories. For both groups (and the two sub-groups of bilinguals), we

Table 3. Multiple regression analyses predicting Cue Impact in the Categorization task. * $p < .05$ for the full model.

	licit Word Impact*		illicit Word Impact*		frame Impact	
Variable	β	p	β	p	β	p
Group (Monolingual, Bilingual)	.086	.500	.333	**.009**	−.044	.748
Linguistic measures						
English Lexical Decision (ACC)	.411	**.003**	.295	**.027**	.178	.220
English Lexical Decision (RTs)	−.272	**.048**	−.037	.778	−.130	.373
Cognitive measures						
Verbal WM	.162	.196	.137	.259	−.171	.206
Visual-Spatial WM	−.191	.122	−.230	.058	−.021	.876
Inhibitory Control	.115	.351	.104	.388	.155	.245
Cognitive Flexibility	−.084	.511	−.243	.054	−.235	.925

were interested in examining whether licit words would have an advantage over other cues after ensuring that all cues were equally discriminable. We were also interested in assessing the contributions of characteristics such as language experience/bilingualism, linguistic skills, and cognitive abilities to the use of these different cue types to extract category boundaries.

Language and categorization in monolinguals and bilinguals

A first major finding from the present work is that language is unique for identifying novel categories for both monolingual and bilingual adults: licit words shifted categorization away from the patterns found in the No Cue control or non-linguistic frames, particularly in the absence of perceptual evidence for the presence of a category (i.e., for standard-target pairs that were only 10% or 30% similar to each other) – even though licit words did not override perceptual similarity (cf. also Gelman & Markman, 1987; Gelman & Davidson, 2013; Diesendruck & Peretz, 2013).

This result is in line with previous developmental studies showing a label advantage in forming novel categories (e.g., Fulkerson & Waxman, 2007), and adult studies showing that words are more effective than non-linguistic cues in activating existing categories (e.g., Boutonnet & Lupyan, 2015). The present finding goes beyond earlier work by showing that labels help adults identify novel categories (not simply activate existing ones) more than non-linguistic cues, and this happens even when labels are equated to these other types of cues in terms of discriminability or intentional presentation.

The present finding also extends prior work by showing that the strong relationship between language and categorization is largely unaffected by the acquisition of a second language. Unlike prior literature suggesting that bilingual learners are more flexible in mapping word forms to meanings (Davidson et al., 1997), and acquire novel words better than monolinguals as adults (Bartolotti et al., 2011; Bartolotti & Marian, 2012; Cenoz & Valencia, 1994; Sanz, 2000; Cenoz, 2003; Kaushanskaya & Marian, 2009; Keshavarz & Astaneh, 2004; van Hell & Mahn, 1997), being bilingual did not affect the likelihood of acquiring unfamiliar but phonologically well-formed words and using them as an aid to categorization. Furthermore, within the bilingual group, length of exposure to a second language did not affect the use of words (or other types of cue) for categorization. Our results are different from prior work on how monolinguals and bilinguals (including different sub-groups of bilinguals) acquire meanings for novel phonological forms (cf. Kaushanskaya & Marian, 2007, 2009). We return to reasons for the similarities between monolinguals and bilinguals, and between the two bilingual groups, in our study in a later section.

A second related finding from this work is that, across both monolinguals and bilinguals, phonologically unconventional, illicit words caused shifts in categorization performance compared to the frame and no cue conditions. Furthermore, even though no differences emerged between monolinguals and bilinguals when the groups were directly compared to each other, bilingualism was associated with use of illicit words for purposes of categorization. This last finding is reminiscent of detailed analyses within the bilingual group showing that bilinguals were just as likely to make use of the phonologically impossible illicit words as the licit words in categorizing novel artifacts. This pattern is selective, since it did not extend to completely non-linguistic cues (patterned frames).

The role of illicit words in categorization, especially for bilingual individuals, is quite striking in light of neuroimaging work indicating that strings of consonants are processed differently from both real words and phonologically possible novel words (Petersen et al., 1990; Dehaene, 1995). This pattern suggests that participants, and especially bilingual participants, were flexible in what they considered a label, sometimes using even impossible strings of letters without any vowels as potential linguistic strings and hence category markers. The association with bilingualism is in line with research showing that bilinguals are at an advantage over monolinguals in acquiring non-traditional form-meaning mappings such as novel words with phonological properties unattested in either of the bilinguals' languages (Kaushanskaya & Marian, 2009) or Morse code languages (Bartolotti & Marian, 2012).

The fact that bilingualism predicts the use of illicit (but not licit) words to draw category boundaries invites more specific explanations. A first possibility is this finding results from bilinguals' increased memory-storage capacity relative to monolinguals: perhaps bilinguals were simply better at the demanding task of remembering consonant strings such as *gsz* (see Papagno & Vallar, 1995). This explanation seems unlikely in the present study, however, given that bilinguals performed similarly to the monolingual group on the digit-span task. Thus the association between bilingualism and illicit word use cannot be accounted for by verbal working memory differences between bilinguals and monolinguals.

A second more likely explanation relates to the extended period for non-native phonological distinctions in bilingual learners (see also Kaushanskaya & Marian, 2009, on the bilingual word-learning advantage). Early exposure to two phonological systems may delay the onset of language-specific phonological tuning (see, e.g., Bosch & Sebastián-Gallés, 2001), and a more flexible phonological system that persists into adulthood may prepare bilinguals especially well for encoding unfamiliar phonological strings and further using such strings as cues for grouping objects into categories.

Language, cognition and individual variation in categorization

A key goal of the present work was to examine the role of individual differences in how language and other cues bear on categorization. We found that, regardless of bilingual language status, variation in dominant language (English) processing skills affected the relationship between language and categorization: individuals with higher English processing skills as measured by a lexical decision task were more likely to be influenced by both licit and illicit words in the categorization task. We also found that variation in executive function capacity did not explain the use of either linguistic or non-linguistic cues in categorization. This last result may appear surprising given the tight link between executive function and categorization (categorization requires keeping relevant properties in mind and inhibiting irrelevant items, and thus categorization tasks are often used as EF measures; e.g., Kongs et al., 2000; Zelazo, 2006). It is possible that our categorization task was not sufficiently cognitively demanding and our participant group (young college-aged adults at peak cognitive performance) not diverse enough for this relationship to emerge.

The finding that individuals with better linguistic skills weigh novel words (regardless of phonological well-formedness) more heavily in category identification than those with poorer linguistic skills meshes nicely with infant data that have shown a link between the use of labels in categorization and the size of the learner's vocabulary (Balaban & Waxman, 1997; Nazzi & Gopnik, 2001; Ferguson et al., 2015). The direction of this relationship remains an open topic of discussion in the literature. In our present data with adults, there are several possible explanations for the link between language processing ability and word use in categorization. One possibility is that better linguistic skills lead some individuals to develop stronger links between novel linguistic labels and object categories. Alternatively, it could be that those individuals who are better at connecting novel labels to objects have an advantage at processing individual words (as measured by the lexical decision task). Yet a third possibility is that both highly precise relations between labels and objects and higher linguistic skills could be attributed to a third variable, such as the ability to identify and remember specific words.

Adjudicating among these three possibilities would require the use of longitudinal studies and larger-scale studies of individual differences. Notice that our current measure of linguistic skills, the lexical decision task, is specifically viewed as a measure of lexical access and was chosen because it has been shown to accurately predict language processing skills in various groups of monolinguals and bilinguals (Harrington, 2006). However, a broader investigation should include additional measures of linguistic skills, including vocabulary size, verbal fluency, picture naming, and sentence comprehension. We should also point out that, in our data, we found no difference between bilinguals' and monolinguals' performance on the lexical decision task, unlike previous work reporting a disadvantage for bilinguals as compared to monolinguals (e.g., Soares & Grosjean, 1984), presumably because our bilingual participants were very proficient in English (the language of the task).

Lastly it is worth pointing out that our data reveal no advantage for bilingual participants over monolingual participants on any of the cognitive tasks in our battery (in fact, on the Wisconsin Card Sorting Task of cognitive flexibility, the bilinguals performed significantly worse than the monolinguals). This is perhaps unsurprising, given the mixed results on the bilingual advantage in the literature (Paap & Greenberg, 2013) and the fact that many other factors besides bilingualism affect executive function (including socio-economic status, physical exercise, and musical training; Valian, 2015).

Limitations and extensions

Our current experiment had several limitations that can be addressed in future work. First, as mentioned already, the present data revealed no differences between monolingual and bilingual groups (or between bilingual sub-groups) in direct group comparisons. However, both sub-groups of bilingual speakers (early Spanish–English and late English–Spanish bilinguals) were English-dominant and were comparable to English monolinguals (and to each other) in terms of both their self-rated proficiency in English and their performance on the English lexical decision task. Since the licit words were based on English phonology (and even the illicit words were phonologically disallowed in both English and Spanish), the homogeneity of the groups in terms of English proficiency may have prevented effects of bilingualism (and sub-types of bilinguals) to surface. A possible extension of the current design would be to explicitly manipulate licit words so that they align phonologically with either English or the bilingual participants' other language (Spanish). Manipulating how English-like or Spanish-like the words are might affect an individual's likelihood to use them, with bilinguals potentially being more likely to make use of licit words that are more Spanish-like compared to English speakers. This new manipulation could also be used to compare Spanish–English and English–Spanish bilinguals, who differed in terms of Spanish proficiency in our study, at least according to self-ratings. (A lexical decision task in Spanish – which can also be used to measure proficiency in heritage speakers and L2 learners (Fairclough, 2014) – could be added in a future version). One possibility is that the higher Spanish proficiency group of Spanish–English speakers would be more likely to rely on Spanish-like labels compared to the lower Spanish proficiency group of English–Spanish speakers. Alternatively, both kinds of bilinguals might be equally likely to make categorization decisions based on any novel word that follows phonological well-formedness rules in at least one of their languages.

Second, future work on the effects of bilingualism in the domain of language and categorization should seek to test speakers with different first and second languages to determine whether a bilingual's specific type of linguistic experience affects the language-cognition interface. For example, native speakers of a logographic language like Mandarin might have been more willing to use cues that native English speakers would interpret as completely non-linguistic (e.g., novel iconic symbols) in a categorization paradigm similar to our own. Future paradigms should also include stimuli other than artifacts that have been argued to be prone to categorization by external cues (such as labels): an obvious candidate would be natural kinds, which tend to be categorized on the basis of 'essences' (Diesendruck & Peretz, 2013; Gelman, 2013).

Third, the present investigation has treated linguistic labels as a cognitive tool that can offer a computational advantage during category extraction. Given that labels are reliable category markers, it follows that their use during categorization can have beneficial effects. However, for other cognitive tasks, using labels is a less helpful strategy, and could even be detrimental. For instance, recruiting language to produce the name of a target object can improve performance in a visual search task but the same self-directed speech impairs search accuracy for less familiar items (Lupyan & Swingley, 2012). Similarly, overtly labelling familiar objects such as chairs and lamps with their basic-level name decreases later recognition performance compared to cases where the objects are not labelled, especially for typical instances of the categories (Lupyan, 2008). It would be interesting to extend the present approach to such cases using both monolingual and bilingual participants.

Fourth, the present study did not focus on semantic differences between a bilingual's languages and the potential consequences of such differences for categorization (see Athanasopoulos, Bylund, Montero-Melis, Damjanovic, Schartner, Kibbe, Riches & Thierry, 2015); Pavlenko, 1999; Marian & Kauschanskaya, 2004). Available evidence suggests that acquiring L2 distinctions in the domain of gender (Bassetti, 2007; Kurinski & Sera, 2011; Kurinski, Jambor & Sera, 2016) or color (Athanasopoulos, Dering, Wiggett, Kuipers & Thierry, 2010; Özgen & Davies, 2002) affects how people use the corresponding categories during categorization tasks. Furthermore, these effects depend on the length of exposure to a second language (Athanasopoulos et al., 2010). It remains to be seen whether the process of forming form-to-meaning mappings for newly-encountered words and later using these words to categorize might be affected by the overlap between word meanings and existing semantic distinctions in one's language.

Final thoughts

Taken together, the present data support the conclusion that language is unique in its ability to influence novel category formation in both monolingual and bilingual adults: both groups spontaneously use labels to form categories, and ignore equally discriminable external non-linguistic cues. Moreover, linguistic abilities – but not cognitive abilities – and bilingualism contribute to systematic individual variation in the use of different types of labels as a tool for categorization. Our results suggest that previously reported 'diffuse' patterns of compliance with linguistic labels in adults' categorization behavior (e.g., Sloutsky, Kloos & Fisher, 2007; Sloutsky & Fisher, 2012) may stem from systematic individual differences. Thus our findings underscore the importance of considering participant characteristics such as language background, or linguistic and cognitive profile, when investigating the complex relationship between language and cognition.

Acknowledgements. We thank Amelia Weiss for assistance with stimuli creation and data collection. This material is based upon work supported by the National Science Foundation under grant #1632849.

References

Athanasopoulos P, Dering B, Wiggett A, Kuipers JR and Thierry G (2010). Perceptual shift in bilingualism: Brain potentials reveal plasticity in pre-attentive colour perception. *Cognition*, **116**(3), 437–443.

Athanasopoulos P, Bylund E, Montero-Melis G, Damjanovic L, Schartner A, Kibbe A, Riches N and Thierry G (2015). Two languages, two minds: Flexible cognitive processing driven by language of operation. *Psychological Science*, **26**(4), 518–526.

Balaban MT and Waxman SR (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology*, **64**, 3–26.

Bartolotti J and Marian V (2012). Language learning and control in monolinguals and bilinguals. *Cognitive Science*, **36**(6), 1129–1147.

Bartolotti J, Marian V, Schroeder SR and Shook A (2011). Bilingualism and inhibitory control influence statistical learning of novel word forms. *Frontiers in Psychology*, **2**(324), 1–10.

Bassetti B (2007). Bilingualism and thought: Grammatical gender and concepts of objects in Italian–German bilingual children. *International Journal of Bilingualism*, **11**(3), 251–273.

Bialystok E (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development*, **70**(3), 636–644.

Bialystok E, Barac R, Blaye A and Poulin-Dubois D (2010). Word mapping and executive functioning in young monolingual and bilingual children. *Journal of Cognition and Development*, **11**(4), 485–508.

Booth AE and Waxman SR (2002). Object names and object functions serve as cues to categories for infants. *Developmental Psychology*, **38**(6), 948–957.

Bosch L and Sebastián-Gallés N (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, **2**, 29–49.

Boutonnet B and Lupyan G (2015). Words jump-start vision: A label advantage in object recognition. *The Journal of Neuroscience*, **35**(25), 9329–9335.

Bull R and Scerif G (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology*, **19**(3), 273–293.

Buchsbaum BR, Greer S, Chang WL and Berman KF (2005). Meta-analysis of neuroimaging studies of the Wisconsin card-sorting task and component processes. *Human Brain Mapping*, **25**(1), 35–45.

Byers-Heinlein K and Werker JF (2009). Monolingual, bilingual, trilingual: infants' language experience influences the development of a word-learning heuristic. *Developmental Science*, **12**(5), 815–823.

Carlesimo GA, Fadda L, Lorusso S and Caltagirone C (1994). Verbal and spatial memory spans in Alzheimer's and multi-infarct dementia. *Acta Neurologica Scandinavica*, **89**(2), 132–138.

Carlson SM, Moses LJ and Claxton LJ (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, **87**(4), 299–319.

Cenoz J (2003). The additive effect of bilingualism on third language acquisition: a review. *International Journal of Bilingualism*, **7**(1), 71–87.

Cenoz J and Valencia JF (1994). Additive trilingualism: evidence from the Basque country. *Applied Psycholinguistics*, **15**(02), 195–207.

Chey J, Lee J, Kim YS, Kwon SM and Shin YM (2002). Spatial working memory span, delayed response and executive function in schizophrenia. *Psychiatry Research*, **110**(3), 259–271.

Daneman M and Carpenter PA (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, **19**(4), 450–466.

Davidson D, Jergovic D, Imami Z and Theodos V (1997). Monolingual and bilingual children's use of the mutual exclusivity constraint. *Journal of Child Language*, **24**(01), 3–24.

Davidson NS and Gelman SA (1990). Inductions from novel categories: The role of language and conceptual structure. *Cognitive Development*, **5**(2), 151–176.

Davidson D and Tell D (2005). Monolingual and bilingual children's use of mutual exclusivity in the naming of whole objects. *Journal of Experimental Child Psychology*, **92**(1), 25–45.

De Abreu PMJE, Gathercole SE and Martin R (2011). Disentangling the relationship between working memory and language: The roles of short-term storage and cognitive control. *Learning and Individual Differences*, **21**(5), 569–574.

Dehaene S (1995). Electrophysiological evidence for category-specific word processing in the normal human brain. *NeuroReport*, **6**(16), 2153–2157.

Deng W and Sloutsky VM (2012). Carrot-eaters and moving heads: Salient features provide greater support for inductive inference than category labels. *Psychological Science*, **23**(2), 178–186.

Diamond A (2013). Executive functions. *Annual Review of Psychology*, **64**, 135.

Diesendruck G and Peretz S (2013). Domain differences in the weights of perceptual and conceptual information in children's categorization. *Developmental Psychology*, **49**(12), 2383–2395.

Eriksen BA and Eriksen CW (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, **16**(1), 143–149.

Fairchild S, Mathis A and Papafragou A (2018). Linguistic cues are privileged over non-linguistic cues in young children's categorization. *Cognitive Development*, **48**, 167–175.

Fairclough M (2014). Spanish as a Heritage Language. In M Lacorte (ed.), *The Routledge Handbook of Hispanic Applied Linguistics*, 134–149. New York/ London: Routledge.

Ferguson B, Havy M and Waxman SR (2015). The precision of 12-month-old infants' link between language and categorization predicts vocabulary size at 12 and 18 months. *Frontiers in Psychology*, **6**(1319), 1–6.

Ferry AL, Hespos SJ and Waxman SR (2010). Categorization in 3- and 4-month-old infants: an advantage of words over tones. *Child Development*, **81**(2), 472–9.

Ferry AL, Hespos SJ and Waxman SR (2013). Nonhuman primate vocalizations support categorization in very young human infants. *Proceedings of the National Academy of Sciences of the United States of America*, **110** (38), 15231–15235.

Fischer MH (2001). Probing spatial working memory with the Corsi blocks task. *Brain and Cognition*, **45**(2), 143–154.

Frank I and Poulin-Dubois D (2002). Young monolingual and bilingual children's responses to violation of the Mutual Exclusivity Principle. *International Journal of Bilingualism*, **6**(2), 125–146.

Friedman NP, Miyake A, Young SE, DeFries JC, Corley RP and Hewitt JK (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, **137**(2), 201–225.

Fulkerson AL and Waxman SR (2007). Words (but not tones) facilitate object categorization: Evidence from 6-and 12-month-olds. *Cognition*, **105**(1), 218–228.

Gelman SA and Coley JD (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology*, **26**(5), 796–804.

Gelman SA and Davidson NS (2013). Conceptual influences on category-based induction. *Cognitive Psychology*, **66**(3), 327–353.

Gelman SA and Markman EM (1986). Categories and induction in young children. *Cognition*, **23**(3), 183–209.

Gelman SA and Markman EM (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development*, **58** (6), 1532–1541.

Gelman SA and O'Reilly AW (1988). Children's inductive inferences within superordinate categories: The role of language and category structure. *Child Development*, **59**(4), 876–887.

Gelman SA (2013). Artifacts and essentialism. *Review of Philosophy and Psychology*, **4**(3), 449–463.

Gershkoff-Stowe L and Smith LB (2004). Shape and the first hundred nouns. *Child Development*, **75**(4), 1098–1114.

Gervits F, Johanson M and Papafragou A (2016). Intentionality and the role of labels in categorization. *Proceedings from the 38th Annual Meeting of the Cognitive Science Society*, 1146–1151.

Gleitman L and Papafragou A (2016). New perspectives on language and thought. In K Holyoak and R Morrison (eds.), *Cambridge Handbook of Thinking and Reasoning*. New York: Oxford University Press.

Gold JM, Carpenter C, Randolph C, Goldberg TE and Weinberger DR (1997). Auditory working memory and Wisconsin Card Sorting Test performance in schizophrenia. *Archives of General Psychiatry*, **54**(2), 159–165.

Grosjean F (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, **36**(1), 3–15.

Harrington M (2006). The lexical decision task as a measure of L2 lexical proficiency. *EUROSLA Yearbook*, **6**(1), 147–168.

Houston-Price C, Caloghiris Z and Raviglione E (2010). Language experience shapes the development of the mutual exclusivity bias. *Infancy*, **15** (2), 125–150.

Johanson M and Papafragou A (2016). The influence of labels and facts on children's and adults' categorization. *Journal of Experimental Child Psychology*, **144**, 130–151.

Jones SS, Smith LB and Landau B (1991). Object properties and knowledge in early lexical learning. *Child Development*, **62**(3), 499–516.

Kane MJ and Engle RW (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, **9**(4), 637–671.

Kaushanskaya M and Marian V (2007). Age-of-acquisition effects in the development of a bilingual advantage for word learning. *Proceedings of the Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Kaushanskaya M and Marian V (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, **16**(4), 705–710.

Keshavarz MH and Astaneh H (2004). The impact of bilinguality on the learning of English vocabulary as a foreign language (L3). *International Journal of Bilingual Education and Bilingualism*, **7**(4), 295–302.

Kessels RP, Van Zandvoort MJ, Postma A, Kappelle LJ and De Haan EH (2000). The Corsi block-tapping task: standardization and normative data. *Applied Neuropsychology*, **7**(4), 252–258.

King J and Just MA (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, **30**(5), 580–602.

Kongs SK, Thompson LL, Iverson GL and Heaton RK (2000). Wisconsin card sorting test-64 card version (WCST-64). *Odessa, FL: Psychological Assessment Resources*.

Kroll JF, Bobb SC and Hoshino N (2014). Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. *Current Directions in Psychological Science*, **23**(3), 159–163.

Kurinski E and Sera MD (2011). Does learning Spanish grammatical gender change English-speaking adults' categorization of inanimate objects? *Bilingualism: Language and Cognition*, **14**(2), 203–220.

Kurinski E, Jambor E and Sera MD (2016). Spanish grammatical gender: Its effects on categorization in native Hungarian speakers. *International Journal of Bilingualism*, **20**(1), 76–93.

Landau B and Shipley E (2001). Labelling patterns and object naming. *Developmental Science*, **4**(1), 109–118.

Lupyan G (2008). From chair to "chair:" A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General*, **137**(2), 348–369.

Lupyan G and Swingley D (2012). Self-directed speech affects visual search performance. *The Quarterly Journal of Experimental Psychology*, **65**(6), 1068–1085.

Lupyan G and Thompson-Schill SL (2012). The evocative power of words: activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, **141**(1), 170.

Lupyan G, Rakison DH and McClelland JL (2007). Language is not just for talking redundant labels facilitate learning of novel categories. *Psychological Science*, **18**(12), 1077–1083.

Markman EM (1991). The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings. In SA Gelman & JP Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 72–106). Cambridge: Cambridge University Press.

Marian V and Kaushanskaya M (2004). Self-construal and emotion in bicultural bilinguals. *Journal of Memory and Language*, **51**(2), 190–201.

Mathôt S, Schreij D and Theeuwes J (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, **44**(2), 314–324.

Miyake A and Friedman NP (2012). The nature and organization of individual differences in executive functions four general conclusions. *Current Directions in Psychological Science*, **21**(1), 8–14.

Mueller ST and Piper BJ (2014). The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *Journal of Neuroscience Methods*, **222**, 250–259.

Nazzi T and Gopnik A (2001). Linguistic and cognitive abilities in infancy: when does language become a tool for categorization? *Cognition*, **80**(3), B11–B20.

Oakes LM and Rakison DH (2003). Issues in the early development of concepts and categories: An introduction. In LM Oakes and Rakison DH (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 3–23). New York, NY: Oxford University Press.

Özgen E and Davies IRL (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General*, **131**, 477–493.

Paap KR and Greenberg ZI (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, **66**(2), 232–258.

Papagno C and Vallar G (1995). Verbal short -term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology*, **48A**, 98–107.

Pavlenko A (1999). New approaches to concepts in bilingual memory. *Bilingualism: Language and Cognition*, **2**, 209–230.

Petersen SE, Fox PT, Snyder AZ and Raichle ME (1990). Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. *Science*, **249**(4972), 1041–1045.

Plunkett K, Hu JF and Cohen LB (2008). Labels can override perceptual categories in early infancy. *Cognition*, **106**(2), 665–681.

Rastle K, Harrington J and Coltheart M (2002). 358,534 nonwords: The ARC Nonword Database. *Quarterly Journal of Experimental Psychology*, **55A**, 1339–1362.

Rhodes M and Gelman SA (2009). Five-year-olds' beliefs about the discreteness of category boundaries for animals and artifacts. *Psychonomic Bulletin & Review*, **16**(5), 920–924.

Rhodes MG (2004). Age-related differences in performance on the Wisconsin card sorting test: a meta-analytic review. *Psychology and Aging*, **19**(3), 482–494.

Romine CB, Lee D, Wolfe ME, Homack S, George C and Riccio CA (2004). Wisconsin Card Sorting Test with children: a meta-analytic study of sensitivity and specificity. *Archives of Clinical Neuropsychology*, **19**(8), 1027–1041.

Sanz C (2000). Bilingual education enhances third language acquisition: evidence from Catalonia. *Applied Psycholinguistics*, **21**(01), 23–44.

Sebastián-Gallés N and Bosch L (2002). Building phonotactic knowledge in bilinguals: role of early exposure. *Journal of Experimental Psychology: Human Perception and Performance*, **28**(4), 974.

Sloutsky VM and Fisher AV (2012). Linguistic labels: conceptual markers or object features? *Journal of Experimental Child Psychology*, **111**(1), 65–86.

Sloutsky VM, Kloos H and Fisher AV (2007). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science*, **18**(2), 179–185.

Sloutsky VM, Lo YF and Fisher AV (2001). How much does a shared name make things similar? Linguistic labels, similarity, and the development of inductive inference. *Child Development*, **72**(6), 1695–1709.

Soares C and Grosjean F (1984). Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access. *Memory & Cognition*, **12**(4), 380–386.

Stoffers D, Berendse HW, Deijen JB and Wolters EC (2003). Deficits on Corsi's block-tapping task in early stage Parkinson's disease. *Parkinsonism & Related Disorders*, **10**(2), 107–111.

Stuss DT, Levine B, Alexander MP, Hong J, Palumbo C, Hamer L, Murphy KJ and Izukawa D (2000). Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: Effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia*, **38**(4), 388–402.

Tanner D and Van Hell JG (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, **56**, 289–301.

Tanner D, Inoue K and Osterhout L (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition*, **17**(02), 277–293.

Valian V (2015). Bilingualism and cognition. *Bilingualism: Language and Cognition*, **18**(1), 3–24.

Van Hell JG and Mahn AC (1997). Keyword mnemonics versus rote rehearsal: learning concrete and abstract foreign words by experienced and inexperienced learners. *Language Learning*, **47**(3), 507–546.

Waxman SR and Markow DB (1995). Words as invitations to form categories: Evidence from 12-to 13-month-old infants. *Cognitive Psychology*, **29**(3), 257–302.

Waxman SR (1999). Specifying the scope of 13-month-olds' expectations for novel words. *Cognition*, **70**(3), B35-B50.

Welder AN and Graham SA (2006). Infants' categorization of novel objects with more or less obvious features. *Cognitive Psychology*, **52**(1), 57–91.

Wolff P and Holmes KJ (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, **2**(3), 253–265.

Wöstmann NM, Aichert DS, Costa A, Rubia K, Möller HJ and Ettinger U (2013). Reliability and plasticity of response inhibition and interference control. *Brain and Cognition*, **81**(1), 82–94.

Yamauchi T and Markman AB (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **26**(3), 776.

Zelazo PD (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, **1**(1), 297–301.