The role of labelling and interaction in the development of linguistic category coherence

Ellise Marissa Suffill

PhD Psychology

University of Edinburgh

2019

Abstract

When we perceive the world, we can carve it up in different ways. People divide up objects into categories in similar ways when asked to do so without language. In contrast, when people use language to label groups of objects, significant differences appear across speakers of different languages. How do labels affect the way that we categorize, and sometimes make us carve up the world more similarly to other people? Secondly, what happens to people's categories when speakers of different native languages interact?

Using a joint-task paradigm and non-conventionalized labels (i.e., non-word labels and non-linguistic labels that lack a conventionalized meaning within the mental lexicon), I firstly investigated how labels can affect category formation across people, in and out of interactive settings. I found that exposure to another sorter during categorization affected the similarity of people's categories, both with and without labels. Because of this, I next investigated how similarly individuals sorted objects with and without novel labels, and with and without a context of communication (i.e., sorting for oneself vs. sorting with other people in mind). I found that novel labels only increased category coherence across people when the context was communicative (i.e., the context required participants to sort with coordination of categories in mind). I argue that this is because language is strongly tied to communication and, as such, language can be used as a tool that helps people coordinate in communicative contexts. Therefore, in contexts in which we do not need to coordinate, novel labels do not appear to yield benefits for category coherence.

Lastly, I investigated the potential differences in category coherence for interactions between speakers of different native languages (i.e., L1-English and L1-Mandarin/L2-English). Results demonstrated that the effects of category-relevant discussion on category structure and coherence are affected by the status of the speaker, on the basis of whether they are an L1 or L2 speaker of the language. Secondly, they showed that explicit coordination does not always lead to increased category coherence between pairs in L1-L2 dialogues. Achieving coherence in representations of categories can be crucial to successful communication both within and across native speakers of different languages, and labelling and interaction play a key role, alongside context, in the development of this coherence.

Lay Summary

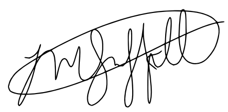
When we perceive the world, we can carve it up in different ways. But having a similar view of the world - and how to label objects within it - is crucial to successful communication. People divide objects into categories in similar ways when asked to do so without language, but when people use language to label groups of objects, significant differences appear across speakers of different languages. What purpose does the act of labelling objects serve, and what factors affect whether labels increase the similarity of people's categories (i.e., category coherence)?

I utilized a joint categorization task that had participants form categories with and without novel labels (i.e., labels with no prior meaning; e.g., non-words), in and out of interactive settings. I found that exposure to another sorter during categorization affected the similarity of people's categories, both with and without labels. Because of this, I investigated how similarly individuals sorted objects with and without novel labels, and with and without a context of communication (i.e., sorting for oneself vs. sorting with other people in mind). I found that novel labels increased the similarity of people's categories specifically when people sorted with a communicative context, and not when they sorted without a communicative context.

Lastly, I investigated the effects of interaction on the similarity of people's categories, when those people speak different native languages and, therefore, have very different ways of labelling objects. I found evidence that interacting about categories might only change the way people categorize when they discuss categories in their native language, and also that, in native-to-non-native interactions, interaction about categories is not enough to bring speakers' categories closer together, even when speakers have the goal of sorting in more similar ways to one another. As such, the need for coordination between speakers may need to be further foregrounded, in order to promote the development of shared understanding between speakers. Achieving coherence in how we label objects can be crucial to successful communication both within and across native speakers of different languages, which is particularly relevant in today's ever-more multilingual world.

Declaration

I declare that this thesis has been composed by myself and that the research reported here has been conducted by myself unless otherwise indicated. This work has not been submitted for any other degree or professional qualification.



Ellise Marissa Suffill

Edinburgh, January 30th, 2019.

Acknowledgements

I'd like to thank the academic and professional staff at the University of Edinburgh who have supported me during my steep learning curve, particularly my three professors: Martin Pickering, Holly Branigan and Antonella Sorace. Thanks to my examiners, Kenny Smith and Larry Barsalou, for their constructive comments and a reasonably brief viva.

Thanks to everyone who made me laugh, or took me for a pint when I needed one. Specifically, thanks to my Edinburgh family: Lauren Hadley, James Dixon and Tattie Hadley-Dixon. Thanks, Michela Bonfieni, for suffering with me as an AThEMEr. Thanks Julie-Anne Meaney for always being down to lose a hat or two on Ilkley Moor. Thanks, Erminia "Bubulandia" Fiorentino, for aborting my numerous office meltdowns and thanks, Rebecca Hellmold, for aborting my numerous flat meltdowns in the Hopeless Crescent. Thanks, Jessica and Anansi Brough for letting me into your lovely (pink) home, and for the never-ending vocal support in the form of miaows. Thanks to Rachael Hulme, who is for life, not just for MSc.

Thank you, fellow Suffills – the two-legged, four-legged and finned varieties – for having put up with me the longest. Special thanks Dad, with his amazing fashion sense, and Mum, for the trademark hair. Thank you to Nanna Poppy and Nanna Milly, for looking after me from baby to adult-baby. Thanks to my patient siblings Natasha (& to my niece Sienna for the programming advice), lil' Melanie and my number one pupper, Kelsey.

Thanks to beautiful, beery Edinburgh for having me (N.B. the wind and Fringe are still terrible). Lastly, thanks Europe, and sorry about Brexit: This research received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613465.

Table of Contents

Chapter 1

Introduction: Concepts, categories and labels

Labels, or the names we give to things in the world, not only reflect the way that we categorize, but also allow us to communicate and share those categories with each other. Labels are thus closely tied to communication. And when we communicate with another person, shared understanding of the world is crucial to the success of that communication. How do labels help us to increase our shared understanding of the world? Do new labels directly influence the way in which we group objects together into categories? Or do we need to learn to use labels in the same way as other people, through exposure to and interaction with those people? What happens when speakers of different languages (who have learnt to label the world in different ways) interact with each other? Does interaction in such cases still increase interlocutors' shared understanding of the world? In this thesis, I will examine the effects of both labelling and interaction on the similarity of people's categories, as a way to examine their shared understanding of the world (i.e., by the way in which they group objects in the world together). Secondly, I will ask whether labels only affect people's categories consistently across people when these labels are used in categorization within coordinative contexts (i.e., contexts in which a label is used to convey information about categories to another person). Lastly, I investigate what happens when speakers with significantly different ways of labelling objects interact with each other, and whether in such cases, these speakers can learn each other's ways of labelling and categorizing everyday objects through interaction.

In this chapter, I will review the existing literature on how our concepts reflect our understanding of the world, and crucially how categories are the external application of these concepts onto objects in the world. From this, I will review current theories and research on how labels not only refer to categories, but can also alter them and, in some cases, make people's categories more similar. In order to predict how novel labels (that lack a conventionalized meaning) might affect categorization, I will then review research focusing on the effects of novel labels on categories. Because labelling can occur in contexts that are coordinative (e.g., interactions such as conversations) and non-coordinative (e.g., non-interactive sorting of objects), I will also review work on how interaction affects categories in relation to labelling. In addition, I will present research on how context might affect both labelling and interaction effects on people's categories. Lastly, I will review research on how having bilingual language knowledge can affect people's labels for categories, in order to better understand how speaking different native languages can lead to people having significantly different ways of categorizing, and what this might mean for the effects of interaction on interlocutors' categories. I will end this chapter with a roadmap of the thesis, summarizing each of the experimental chapters and how I intended to use them to test how different factors affect the way that we categorize and may influence us to sort in more similar ways to one another, with the overall goal of discovering their potential benefits to understanding and communication across speakers.

1.1 Concepts across people

Concepts are abstractions of the rules and dimensions that make up the fundamental features we associate with a certain notion, or object. Take the example of a bottle; whilst different bottles may have different features in terms of size, material, colour and so on, we can extract the generalizable features that encompass the essence of what makes something a 'bottle' and interpret this as the concept of BOTTLE (see Fig. 1). Whilst there is often overlap across people's concepts, each individual may have a slightly different concept than the next, as seen in everyday disagreements about 'what constitutes an X'. For example, people may argue about whether soft plastic containers for water (often called 'water bladders') can also be referred to as 'bottles'. Such disagreements may occur due to peoples' different rules, or dimensions, about the material or structure (e.g., rigidity) of what exactly constitutes BOTTLE.

In most situations people can agree upon the content of most concepts and they do so while referencing these concepts through words (e.g., in conversation). As such, two interlocutors discussing bottles may both have slightly different individual concepts of BOTTLE, but in discussion their representations have enough overlap to allow for successful communication. The commonalities in human biology and perceptual experience may explain why many of our concepts share enough similarity, such that there is sufficient conceptual overlap between people (Rosch & Mervis, 1975; Johnson, 1987). It is when the alignment, or coordination, of people's concepts fail that communication can breakdown and some form of repair has to occur between interlocutors (Pickering & Garrod, 2004).

Fig. 1. Demonstration of how the common features of individual instances of 'bottle' could be abstracted to a more general example of the concept of BOTTLE.

Concepts are subject to change through factors such as interaction and context, meaning that individuals' concepts are not always identical and can be contingent on the current context or situation (Casasanto & Lupyan, 2015). The concept of BOTTLE, for example, would be subject to change across each instantiation. Existing knowledge and experience can affect the types and frequency of information activated when one hears the word 'bottle', but no two instances of the concept or the items it encompasses are guaranteed to be identical (Barsalou, 1983). For example, if a person rehydrating themselves after playing sport is asked 'do you want another bottle?' the representation of BOTTLE activated in their head is likely an image of a plastic bottle filled with water, or some form of energy drink. But if the same individual is drinking at a bar, and is asked 'do you want another bottle?' their representation of BOTTLE is more likely to reflect a glass bottle, filled with their alcoholic beverage of choice. As such, differences in how we perceive, or realize, concepts across different instantiations are contingent on situation-specific factors, or cues-in-context, which can relate to physical and communicative environmental cues, such as where we are and the things we likely to encounter or discuss there.

1.2 Categories as proxies to concepts

Categories are said to reflect concepts, in that categories are essentially the application of concepts to groups of items (i.e., items that pertain to the rules, or dimensions, of a concept), whether those items are physical objects, or abstract notions (Murphy, 2002). Murphy describes concepts as 'mental representations of classes of things' whilst categories 'talk about the classes themselves' (pg. 5). As such, the two are intrinsically linked such that categories describe concepts by identifying the items that fit under a particular concept's rules and dimensions.

Categories, like concepts, allow for and also rely on generalizations about items and objects. If we encounter a novel bottle-like object, it may not be identical to our concept of BOTTLE, but if it shares enough features with our general representation of BOTTLE, then it is likely that this specific object will be categorized alongside other, bottle-like objects we have encountered and previously categorized as bottles (Medin & Schaffer, 1978; Rosch, 1978). If, instead, we treated every object we encountered as completely novel, and had to form an individual representation for each item, the cost of storing each and every representation becomes unfeasible (Rosch & Mervis, 1975). As such, objects are grouped by similarity into categories.

As for the central items and objects that help conceptualise a category, different theories posit different ways in which we attain a good, generalizable idea of what a specific concept entails and what its category is likely to comprise of. *Exemplar theory* suggests that the items most central to a category (i.e., those items which comprise the most common and identifiable features of the category and the underlying concept that defines it) act as exemplars for that category. As such, when deciding how to categorize a novel item, we would compare its features to the features of previously encountered exemplars in order to decide whether the novel item fits into that category (Medin & Schaffer, 1978). In contrast, *prototype theory* suggests that central to each category is a prototype, or a summation of the most typical features common to items from the category in question (i.e., rather than an item actually encountered in the real world; Rosch, 1978). Lastly, *simulation theory* posits that that people have theories about the world that allow them to bind together seemingly unrelated features, and they use these theories as a comparison for whether certain items fit within a certain concept and category, or not

(Murphy & Medin, 1985). Despite differing accounts of how inferences for category membership occur, these theories acknowledge the importance of the link between concepts and categories. As the exact nature of categorization is not the focus of this thesis, I will now focus on how categories, as a reflection of concepts, may change due to labels and interaction.

It should be noted that the aforementioned theories of classification seem to lend themselves to the explanation of more concrete concepts and, thus, the categorization of physical items. In contrast, abstract concepts refer to entities that are not physical or spatially constrained (e.g., particular feelings or experiences; Barsalou & Wiemer-Hastings, 2005). Therefore, the items which would be categorized under abstract concepts are themselves either abstract, non-physical things (e.g., the feeling of love), or more tangible, physical objects, but those which are linked to the concept in more abstract ways (e.g., roses relating to and being categorized under the concept of love, as this is a context in which roses commonly occur in Western cultures). These more abstract types of concept and category can be difficult to represent and measure in a tangible manner, since the links between objects within the category tend to be much more varied in their perceptual and functional features (i.e., compared to a concrete category, like BOTTLE). Because of the nature of abstract concepts, it is therefore easier to measure category change and similarity across people by focusing on concrete concepts and categories for real world objects.

In this way, categories have been used by researchers as proxies for more concrete concepts: that is, by studying people's categories, we can infer information about what an individual's concept of a certain thing entails (Murphy, 2002). For example, if a person asked to categorize container-like items only ever puts items with a long, slender neck in the 'bottle' category, then we can infer that the person's current concept of BOTTLE has a feature, or dimension, reflecting neck length (i.e., items under the concept of BOTTLE, and in the category of 'bottle', generally comprise relatively long necks, compared to other container-like objects). We can also use categories to examine changes in people's concepts (i.e., via changes in their categories). The work in this thesis will examine how labelling categories with words and interacting with another person about categories may cause such category changes. Specifically, I will examine whether these factors can increase category coherence between people, and as such I will next address the literature on how words – as labels – can affect people's categories.

2.1 Concepts, categories and words as labels

When we interact about objects in the world with another person, we can use words to label the things that we are discussing (e.g., ‘Can you pass me that bottle?’). As such, words and categories occur together in the world whenever we use language to describe or refer to things around us (Laskowski, 2010). In such cases, we inherently label items as belonging to one category or another. Some researchers argue that while categories may be the external application of concepts to real world items (Murphy, 2002), words can directly label both the concept itself and the category (i.e., the items that fall within the category; Laskowski, 2010). As such, words as labels share an important link with both concepts and categories. When we group together objects by using a word to label them, this forms a linguistic category (i.e., all objects we would call ‘bottle’ are under the linguistic category of ‘bottle’).

However, we can also think about things without naming them. For example, if a person is sorting the recycling, they may sort bottles and tins into separate bins, without ever having to explicitly call into mind the word labels for these categories. As such, we can categorize objects together without necessarily needing to apply a word label to those categories, and to objects within those categories. Because of this difference between labelled and non-labelled categories, categorization can take on two forms: linguistic versus non-linguistic categorization. Linguistic categories are those formed of groups of items that share the same name, or label, whereas non-linguistic categories can be thought of in terms of how we categorize items without having to apply word labels to them (i.e., how we sort items in the absence of linguistic labels).

2.2 Non-linguistic categories

Perceptual categories are formed on the basis of visual perceptual features, including object shape, colour and size. Perceptual categories are acquired by pre-linguistic infants, with evidence for the existence of perceptual category distinctions for some natural kinds (e.g., cats and dogs) in infants as early as 3- and 4-months old (Quinn, Eimas & Rosenkrantz, 1993). In fact, infants, great apes, monkeys, rats, and birds can all learn a variety of basic level perceptual categories (Cook & Smith, 2006; Quinn et al., 1993; Smith, Redford, & Haas, 2008; Zentall, Wasserman, Lazareva, Thompson & Rattermann, 2008) which provides support for the notion that perceptual input provides the basis for the development of concepts in infants (i.e., in contrast to nativist accounts of innate concepts; cf. Chomsky, 1980, Gelman, 1990).

Sloutsky (2010) argued that infants initially learn dense perceptual categories (i.e., categories in which items have multiple inter-correlated features for membership, as is often found with natural kinds like CAT or DOG) before they learn to integrate cross-modal information. In contrast, sparse categories are those in which items share only one or two features for category membership (e.g., mathematical or scientific concepts, such as GROWTH in an animal population, and GROWTH in a bank account's interest). Integration of cross-modal information eventually allows some categories to be lexicalized, through the infant learning to associate visual (perceptual) input with auditory (lexical) input. Once children begin to learn dimensional words for values (like 'red' or 'square'), they can start to acquire more sparse categories (in contrast to dense categories, like 'cat' or 'dog'). Eventually, the learning of sparse categories supports the development of abstract concepts, with evidence that in some cases, children learn the word (or label) for abstract concepts, as a precursor to learning what that abstract concept actually entails (Vygotsky, 1964).

Even in adults, perceptual categories tend not vary across cultures and across speakers of different languages (Malt, Sloman, Gennari, Shi & Wang, 1999). They are thought of as 'natural categories' with highly pre-determined courses of development across humans (Rosch, 1973; Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976). These types of perceptual categories are argued to not vary significantly across people because perceptual categories are based upon physical features that we perceive. The way humans process and understand perceptual features tends to be similar across all humans because we share the same physiological basis for perception. Thus, because people share perceptual experience of the same structured world, their perceptual categories (as well as the related concepts) tend to overlap (Rosch & Mervis, 1975; Johnson, 1987). There are of course some exceptions to this, in cases where perceptual categories have been heavily influenced by language and culture, such as colour categories (e.g., Roberson, Davies & Davidoff, 2000; Thierry, Athanasopoulos, Wiggett, Deringa & Kuipers, 2009). Overall, however, perceptual categories are relatively stable across humans, since they are based on physical, visual features that we generally – as a species – experience in the same way.

Another form of non-linguistic category is the functional category. Functional categories are based upon the function that an object serves, or is used for, and thus objects are grouped together on the basis that they share a similar function (e.g., items that are used to hold beverages). Object functions can also draw infants' attention to perceptible features and, thus, aid categorization of objects that share such features by around 14-months of age (Waxman & Booth, 2001). Infants can use shared function to correctly select new category members for a familiar category, and rely on object function more so than object name until around 18-months old (Booth & Waxman, 2002; Booth, 2006). However, object names have also been shown to facilitate categorization for 14-month-olds, when both the object name and a hint about the function of the object is given (Booth & Waxman, 2002). So, again, it appears that labelling may interact with functional features, as it does with perceptual features.

Function continues to affect categorization in adulthood; adults rely on object function over object shape during the lexical extension of novel labels to new objects, when function is emphasised during learning (Graham, Williams & Huber, 1999). Functional categories appear to show slightly more variation than perceptual categories, across speakers; Malt et al. (1999) attributed this to the fact that different cultures have different practices and, thus, different objects to serve functions in those practices. Another aspect is marketing and branding (e.g., a branded character depicted on a juice box may be recognisable to a person in the US, but not in China, due to that brand not being sold in China), and this may influence the way in which the two cultural groups would categorize branded objects (Malt et al., 1999). However, Malt et al. argued that the variation across cultures (and so, speakers) for functional objects is still significantly lower than the variation found in linguistic categories for dishware (e.g., 'bowls', 'plates', 'jars', etc.) across speakers of different languages.

2.3 Linguistic categories

There exists an obvious underlying perceptual and functional basis for many linguistic categories, such that objects that have important features in common tend to be given the same name. Despite this close connection, Malt et al. (1999) argued that the act of naming differs significantly from the act of recognition: Naming is part of a communication process, whereas recognition is not. The name selected for an object may reflect the requirements for successful communication, whereas the non-linguistic representation of commonalities is presumably influenced primarily by constraints such as storage efficiency and the ability to support inference. And because of this fundamental difference in the nature of the two acts, there can be inherent differences between linguistic and non-linguistic categories.

Slobin (1987) argued that language is a 'special form of thought that is mobilised for communication' (pg. 436). In contrast, Malt et al. argued that there is a dissociation between language and thought – that is, the dissociation between the labels we use for items versus the categories we would place these items into regardless of linguistic information. Regardless of the exact nature of language in relation to thought, both accounts predict differences in how we think about the world when we do so with language, versus when we do so without using language.

In line with these predictions – and in contrast to perceptual and functional categories – linguistic categories appear to be relatively distinct from non-linguistic categories, in that they show substantial cross-linguistic variation (Kronenfeld, Armstrong & Wilmoth, 1985; Malt et al., 1999). Given that sharing information about objects and categories is a core part of everyday communication, and that we share this information by labelling those objects (and categories) through language, it becomes important to understand the cross-linguistic differences in linguistic categories, and why they emerged.

Malt et al. (1999) discussed three forms of complexity in naming that may have led to this dissociation between linguistic and non-linguistic categories. In *chaining*, a target item (T) is semantically situated closest to one semantic cluster (C1, see Fig 2.). However, intermediate semantic links between T and items from a second semantic cluster (C2) may lead to T being given the same name as items from C2, as opposed to the same name as items from C1. Specifically, through chaining T becomes associated more with C2, and is labelled as such, despite the target item having greater perceptual similarity to other items in the original cluster of C1 (Heit, 1992). An example of this could be a plastic juice container with a straw (see Fig. 3) being referred to as a 'juice box', despite it sharing overall more perceptual features with that of a 'bottle'. For example, this chaining could have occurred because such an item developed from more prototypical cardboard juice boxes and so, retains its associations with items labelled as 'juice box', more so than it does with 'bottle'. In this way, the object retains some functional – but not perceptual similarities – with a prototypical 'juice box', and, therefore, is given the label 'juice box', not 'bottle'.

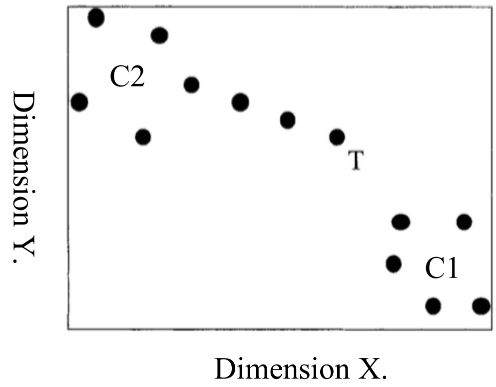


Fig. 2. Example of *chaining* between target item (T) and secondary cluster (C2) within a semantic space (for two hypothetical dimensions). Adapted from Malt et al. (1999).

Fig. 3. Example of an item labelled as 'juice box'. Taken from Malt et al. (1999).

Secondly, through *convention*, simply being told that 'object X is a juice box' can also create strong associations for labelling that object, such that a person relies upon this convention, and ignores X's perceptual or functional similarity to the prototypical objects that we would call 'bottle'. As such, an object may come to have a particular label through linguistic convention (Clark, 1993; Golinkoff, Mervis & Hirsh-Pasek, 1994; Lehrer, 1990), rather than because of its perceptual or functional similarity to other items that have a different label.

Thirdly, in *pre-emption* (Clark, 1988, 1993; Clark, 1991; Lehrer, 1990; Markman & Wachtel, 1988; McCawley, 1978), speakers pre-emptively use a different label for a target object to reduce potential ambiguity with other, perceptually similar items (e.g., referring to something used to serve soup as a 'soup tureen' rather than as a 'soup bowl'). This leads to differences in how the same object would be categorized individually versus when we categorize the object in a specific context in relation to other items (e.g., items that are more prototypically called a 'soup bowl'). In this way, context can pre-emptively influence how we would name an object, this particular situation relating to the function of the object, in a way that may differ from how we would name the object on a more perceptual basis (e.g., since a soup tureen is essentially an oversized soup bowl, sharing many of the same perceptual features aside from size). Overall, these three forms of complexity help to explain how linguistic categories can consistently vary from non-linguistic categories.

In their experiments, Malt et al. (1999) asked people to rate the similarity of pairs of household objects without labels, and computed pairwise similarity ratings (i.e., how similarly each possible pair of participants sorted the items) for speakers of three different languages (i.e., American English, Argentinian Spanish and Chinese Mandarin). People then named each of the objects individually and the similarity for naming patterns was computed across language groups. Whilst there was significant overlap in how people rated the items on perceptual similarity (i.e., without labels), labelling patterns in the naming task differed significantly across speakers of different languages. That is, speakers of one language were more likely to categorize objects (e.g., dishware, containers) differently from speakers of other languages when asked to sort the objects using labels (e.g., 'box', 'carton', 'bottle', 'jar') than when they are asked to sort the objects without using labels. Malt et al. referred to this as a difference in the *knowing* versus the *naming* of objects. This difference in linguistic categories is partly attributable to the fact that different languages use a different number of labels to divide up items. For example, given 60 household objects a speaker of English might use seven labelled categories, whilst a Chinese Mandarin speaker might use only five (Malt et al., 1999).

There can also be differences across languages in where items lie with respect to category boundaries. For example, whilst both English and Chinese speakers might label a cushioned, four-legged object a chair, a similar item for seating multiple people would still be referred to as a chair in Chinese, but as a sofa in American English (Malt, Sloman & Gennari, 2003). These results appear consistent with a shift away from a reliance on primarily more perceptual features when categorizing without language, due to the aforementioned complexities associated with object naming.

However, Laskowski (2010) argued that the stimuli used in Malt et al. (1999) were biased towards a US population, in terms of cultural exposure to the branded objects in the stimuli set. That is, some of the objects in Malt at al. consisted of branded items common in the US at the time of the study (e.g., a Mickey Mouse shaped bottle), but less so in Argentina and China. As the speakers of Argentinian Spanish and Mandarin Chinese were less exposed to these branded objects, this may have reduced their ability to accurately judge items' similarity on the basis of function, for example.

As such, Laskowski replicated the experimental design of Malt et al. (1999), but used only bowls and plates (which included items morphed to contain features of both) as stimuli. Laskowski chose to use a narrower, more ambiguous range of items to increase variance in categories across people (i.e., via the inclusion of items whose category identity was debatable). Under the same design but with less culturally-biased stimuli, Laskowski overall found that, again, people categorized more similarly to one another, regardless of native language, when they sorted without labels than when they sorted items with labels, but also that sorters who spoke the same native language had more similar linguistic categories, than speakers of different native languages. As such, labels may support the coherence of concepts and categories for speakers of the same language.

In this way, we can use language to reinforce categories (and their underlying concepts) through the application of word labels, that not only help us learn category boundaries and increase consistency in sorting, but which may come to affect the way in which we categorize. Labelling might affect categories through top-down effects on processing caused by the existing meaning of a label, and the features of other items already associated with that label. For example, hearing that an object is a 'turtle' drives us to attribute to that object properties we associate with other, previously encountered objects that have also been labelled as 'turtle', properties such as being reptilian and cold-blooded (Gelman, 2003; Gelman & Davidson, 2013). While category induction does not require language (Baldwin, Markman, & Melartin, 1993), giving category labels for groups of items can foster inductive inferences. For example, the presence of a shared label aids infants and young children in the categorization of dissimilar objects (Dewar & Xu, 2009; Waxman, 2004), with adults also showing similar effects of labelling (Lupyan Rakison, & McClelland, 2007).

Labelling can also influence early visual processing during categorization. For example, Lupyan and Spivey (2010) showed that hearing a redundant label (i.e., a label that provides no new information relevant to the task) still effectively guides visual attention towards target objects. In their experiment, hearing ‘five’ more quickly and accurately guided participants’ attention to the 5s (in a display of 2s and 5s) during a visual identification task than was the case on trials in which participants did not hear this redundant label. This effect occurred even though participants knew – prior to hearing the label – that their task was to attend to only the 5s.

Lupyan (2008) suggested the people undergo a representational shift when they use labels to categorize items, and that this shift accounts for how labels can also change how we represent and remember items. Specifically, this account suggests that labelling can affect categorization by causing a shift in how people represent categories, and that it does so by distorting their memory for the physical features of an object most reliably associated with a category label. That is, when speakers apply a category label to an object, it causes the representation of the object to become a mix of its idiosyncratic features and the features typically associated with the relevant category. Consistent with this claim, participants were worse in a recall task at identifying objects of furniture that they had previously seen if they had named the object (e.g., they had identified it using the labels ‘table’ or 'chair') than if they had made a like/dislike decision for it (i.e., without labelling the object). They performed most poorly when recalling previously labelled objects that were highly prototypical and unambiguous (i.e., objects that had the fewest idiosyncratic features to differentiate them from common category features).

Lupyan (2008) argued that this representational shift plays a facilitatory role in categorization, by helping people to select more abstractable category dimensions. The way in which this might occur is that labels, among other things like exposure to others’ categories, could influence people to select perceptual bases for dimensions, instead of more individualistic ones, and these perceptual bases tend not to vary across humans. From this, they are able to avoid forming categories on too numerous and fine a set of dimensions, which would lead to a greater and potentially unhelpful number of categories. The reasoning behind this is that when people use fewer, more abstractable and perceptually-based dimensions, it becomes more likely that their categories will overlap and be similar to each other. This conceptual shift in how people select category dimensions reorganises which category features the sorter prioritises when deciding how to divide up the objects and can increase the similarity, or coherence, of people's categories by driving attention to the same relevant features across sorters (Laskowski, 2010; Lupyan & Thompson-Schill, 2012; Barnhart, Rivera & Robinson, 2018). This might help explain why speakers of the same native language categorize items (like dishware) more similarly to one another when they use labels to name the items, than when they do not (Laskowski, 2010).

2.4 Effects of novel labels on categories

Taken together, this evidence shows that the use of existing, conventional labels influences how people categorize objects and may sometimes bring people's categories closer together. But people frequently encounter novel labels for new or indeed established objects (e.g., names for new technologies), and such novel labels can also influence category learning. These labels can be novel in the sense that they either lack any conventionalized meaning (i.e., they are 'semantically empty', Lupyan et al., 2007), or that they hold some conventionalized meaning, but not in relation to the objects we are currently labelling (i.e., when a new meaning is learnt for an existing word). In both cases, these labels are initially arbitrary in relation to the new objects we are labelling, but the labels can develop meaningful associations to the objects through simultaneous presentation.

Novel, non-word labels can be used to examine the interplay between object dimensions and the meanings that develop for the novel labels applied to them, since using novel words has the advantage of labels not being already tied to specific entries in the mental lexicon (cf., Jackendoff, 2002). This is the case as long as the novel labels do not too strongly reflect existing and meaningful word forms, for example through sound symbolism (Köhler, 1929). This gives greater potential for new label meanings to develop on the basis of the object dimensions that the labels may come to refer to. During the process of association, labels actively influence category formation by guiding attention to the relevant perceptual similarities of objects for both infants (Waxman & Markow, 1995) and adults (Lupyan & Casasanto, 2014). For example, presenting non-word labels (e.g., 'Look at the Timbo!') to infants alongside new objects guides their eye movements to common features across those objects, so that they increasingly direct their attention more towards shared features than towards dissimilar features, and this leads to enhanced category learning (Althaus & Mareschal, 2014; Althaus & Plunkett, 2016).

Adults similarly show effects of non-word labels on category learning. Lupyan and Casasanto (2014) demonstrated that the application of meaningless, non-word labels (i.e., 'fooves' or 'crelches') to novel, 'alien' stimuli can reinforce the categorization process. This categorization was based upon physical characteristics, and the novel cues worked as successfully as the conventional cues ('smooth' or 'pointy') at increasing category learning. Lupyan and Casasanto (2014) argued that the presence of the novel labels drew attention to the necessary perceptual features (i.e., pointedness vs. smoothness) across objects that supported category learning within individual participants.

Their results also showed that across participants, people came to more effectively associate the same stimuli with the same labels for congruent trials (e.g., participants commonly categorized the same, smooth-headed aliens as 'fooves'). This is unsurprising, as participants were learning pre-determined categories with pre-assigned labels (i.e., their task was to learn a set of distinctions, including a distinction between two labels and the objects to which they referred) and although these non-words did not have conventionalized meanings, they were constructed to reflect the category dimensions (i.e., smoothness vs. pointedness) through sound symbolism (Köhler, 1929). Nevertheless, their results demonstrate that under at least some circumstances, labels that do not have a conventionalized meaning within the mental lexicon can also have a consistent effect on how individuals form categories.

Lupyan et al. (2007) argued that the effect of labels is not necessarily restricted to spoken or written language, and that in principle almost anything could act as a category label. For example, although hearing humans tend to use written symbols and sound sequences as labels, deaf people have learnt to use motor gestures instead. Nevertheless, there appear to be some limits to what can serve as advantageous labels (i.e., labels that facilitate category formation). Lupyan et al. (2007) investigated the efficacy of printed and spoken non-word labels (i.e., *leebish* vs. *grecious*) compared with non-linguistic, location-based cues as labels (i.e., location of the object onscreen) in learning to categorize novel alien stimuli as ‘friendly’ or ‘unfriendly’. Location-based cues involved the alien moving vertically onscreen to indicate ‘where it lived’ (e.g., whether the alien came from the ‘friendly’ part of the planet, or not). Both printed and spoken word labels facilitated category learning, but location-based cues did not.

Lupyan et al. (2007) argued that word labels were effective because they simplified the distinction between the categories. That is, using word labels meant that participants could categorize the objects under a single term (here, *leebish* or *grecious*) that represented multiple category dimensions and also made these dimensions more concrete. In contrast, when they did not use labels, they had to rely on more complex and fuzzy perceptual distinctions (e.g., objects categorized as ‘more rounded and smooth’ and ‘less rounded, with ridges’). They argued that location-based cues did not serve as a label for the existing category dimensions, and so did not facilitate category learning, perhaps because adults prioritize words as referring labels (e.g., nouns as object names), but do not commonly use facts in this way (i.e., do not use facts – such as where an alien lives, as indicated by its location cue – as a way to refer to different objects; Colunga & Smith, 2005). Another possibility could be that location-based cues did not facilitate category learning because they did not simplify the distinction between categories; rather, they added a further dimension to a category (e.g., an object became represented as the three dimensions of ‘more rounded, smooth and with an upwards motion’, instead of the location cue coming to reflect the dimensions of ‘more rounded and smooth’).

A further constraint on novel labels that affect categorization is the extent to which they can be sufficiently abstracted to represent the combination of dimensions that make up a category, and not to refer to too small and specific a set of items to usefully serve as a category label. This constraint relates to the aforementioned importance of being able to use abstractable rules and dimensions (such as those based on perceptual features) for categorization, in order to be able to apply these rules to a wider range of items within a category (Lupyan, 2008). In Edmiston and Lupyan (2015), participants heard a cue (i.e., a verbal, linguistic label, or an environmental sound) and were presented with an image of an object. For each category of object (e.g., ‘guitar’), there were two object instances (e.g., an acoustic guitar vs. an electric guitar) and two environmental sound types (e.g., the strum of an acoustic guitar vs. the strum of an electric guitar). Their task was to decide whether an object belonged to the same basic category as that of the cue. For the 50% of trials in which objects did match the cue, the cue-object congruency was randomized across participants. On congruent trials, for example, participants would hear an acoustic guitar strum and see an acoustic guitar, while on incongruent trials, participants heard an acoustic guitar strum but saw an electric guitar. As such, hearing an acoustic guitar strum cue, but seeing an electric guitar would still require the participant to respond that the object came from the same basic category as the cue (i.e., 'guitar').

For both congruent and incongruent trials, verbal linguistic cues elicited faster response times than did environmental sounds. On the environmental sound trials, participants verified images that better matched the likely source of the sound (i.e., congruent sound trials) more quickly than on incongruent sound trials. Critically, when cued by a label, participants responded faster, even compared with specifically the sound congruent trials. Edmiston and Lupyan, therefore, argued that linguistic labels (such as 'guitar' or 'dog') are advantageous for processing category members because they are superordinate (i.e., they can be abstracted over all kinds of dog or guitar). In contrast, environmental sounds (such as a dog's bark, or the strum of a guitar) do not produce the same advantages in categorization because they reference specific instances of objects (Lupyan & Thompson-Schill, 2012). For example, a high-pitched yap plausibly only refers to instances of small dogs like Chihuahuas, but not Border Collies or Great Danes. In contrast, the word 'dog' can refer to all of these instances. In conclusion, Edmiston and Lupyan argued that labels are special, because they activate concepts in a more abstract, decontextualized way, such that labels act as 'unmotivated cues' to categories. I argue, in line with Lupyan et al. (2007), that the label need not be linguistic (i.e., a word), as long as it can serve the same purpose.

Evidence so far supports the existence of labelling effects upon individuals in non-interactive settings. However, labels are most often used to refer to objects (and, thus categories) when interacting with other people. Given that interaction between people is also posited to bring their linguistic representations closer to one another through the process of alignment (Pickering & Garrod, 2004), it is possible that labelling effects may interplay with the effects of interaction. As such, I will now discuss the effects of interaction on the coherence of people's linguistic categories, and then I will compare this with cases of non-interactive alignment, or coordination that is not based upon interaction between people.

3.1 Interactive alignment

Having a shared understanding of the world can be crucial for successful communication, and this can depend on the alignment of representations between interlocutors at different linguistic levels (Pickering & Garrod, 2004; 2006). How then, do people reach a shared understanding in dialogue? While private, internal concepts and public, shared concepts may not be identical, interlocutors still need to align on their understanding of a given concept and its category enough to allow for successful interaction. *Alignment* refers to formation of these shared representations between interlocutors (Pickering & Garrod, 2004; 2006). Linguistic levels of representation include the conceptual/message level, syntactic level, lemma/word level, and finally the levels relating to the articulation of the message through aspects such as phonology and prosody (Pickering & Garrod, 2013). Levelt's (1992) model of language production (see Fig. 4) forms a basis for defining the levels involved during production. As Pickering and Garrod's account of alignment tightly couples production and comprehension (in order to allow for an imitative account of alignment, through the internal simulation of a speaker's utterance by the listener), aspects of the production system can also be considered active during comprehension.

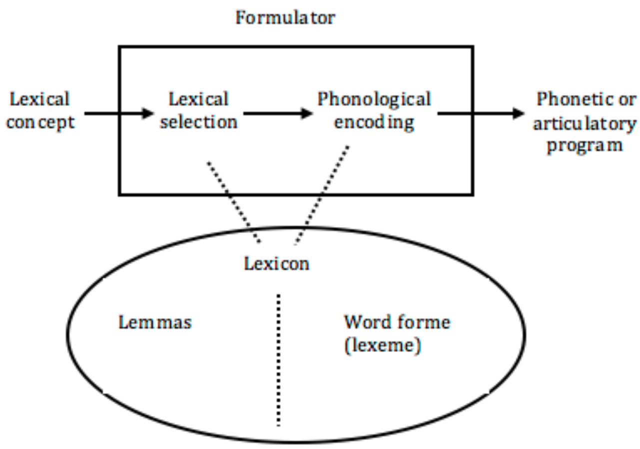


Fig. 4. Levelt's (1992) model of language production.

While Levelt's model relates to a feed-forward account, Dell and O'Seaghdha (1992) argue that the flow of activation across the system is not only a feed-forward mechanism (e.g., from concepts down to articulation), but that it is instead bi-directional. In this way, all levels of representation (i.e., semantic, syntactic, lexical and phonetic) may provide feedback to each other. A bi-directional approach could account for self-monitoring and correction within dialogue (which is another important component of Pickering & Garrod's alignment account), unlike feed-forward only accounts. Given this, it is intuitive to consider alignment as a process which can occur both within and across several levels of linguistic representation, in both directions, and indeed there is evidence to suggest that this is the case (e.g., Branigan, Pickering & Cleland, 2000; Cleland & Pickering, 2003; Schoonbaert, Hartsuiker & Pickering, 2007).

Pickering and Garrod (2004) argue that interactive alignment simplifies language processing in dialogue by: (a) supporting a straightforward interactive inference mechanism, (b) allowing the development and use of routine expressions, and (c) supporting a system of monitoring that can make the repair of misalignments easier. Because the resulting overlap between interlocutors' representations, they argue that conversation in general does not require the costly routine modelling of an interactive partner's mind. Instead, the overlap in representations is sufficient to allow for specific contributions by a speaker to trigger changes in the listener's representation, and/or to bring about interactive repair between interlocutors. This reflects the notion that 'the listener will retain an appropriate model of the speaker's mind, because, in all essential respects, it is the listener's representation as well' (pg. 180).

In comparison, processing speech from monologue is more costly than dialogue, because it lacks automatic alignment and interactive repair between its participants. As such, the listener must resort to bridging inferences when they fail to understand something, while the speaker must engage more explicit mechanisms of audience design to try and avoid such failures in understanding (Clark & Murphy, 1982). Schober's (1993) findings in spatial perspective taking support this notion such that speakers in monologue were more likely to adopt a 'costly' listener-directed frame of reference (e.g., 'on your left'), than were speakers in dialogue (who tended to retain more egocentric representations). This listener-directed approach is more costly, than the egocentric approach commonly found in dialogue, as the former requires the speaker's modelling of the listener's perspective and inhibition of their own perspective (Schober, 1993;

Pickering & Garrod, 2004). In contrast, the use of more egocentric representations in dialogue was effective presumably because alignment allowed both partners to more easily understand each other's perspectives by having shared representations. In summary, speakers took different perspectives (i.e., egocentric vs. allocentric) when they performed the task with a conversational partner in dialogue, as compared with when they performed the task in monologue.

Pickering and Garrod (2004) argue that alignment can be explained through a process of priming. Under this approach, production and comprehension are tightly coupled (Pickering & Garrod, 2007; 2013). That is, when we comprehend the our interlocutor's production, we internally and covertly simulate their action; or in this scenario their utterance, leading to the co-activation of associated structures and concepts within ourselves. Thus, alignment could be seen to work via priming mechanisms between interlocutors. They argue that this primitive priming mechanism does not require additional processing effort and does not rely on explicit negotiation between interlocutors. Through priming, interlocutors do not need to model each other's mental states and instead simply align their representations, and repair misalignments via adjustments to the situation model. They can also make adjustments to new input so that input fits with both interlocutors' existing models. This process can be seen in the case of a dialogue excerpt from Garrod and Anderson's (1987) maze studies, in which partners adjust their input in order to fit with the shared model and avoid misunderstandings about their own/their partner's location within a grid-like maze (here, B formulates a response to fit with the model put forward by A):

A: You are starting from the left, you're one along, one up? (2 sec.)

B: Two along: I'm not in the first box, I'm in the second box.

As for how priming and, thus, alignment works across different levels of representation (i.e., rather than, say, just within syntactic representations, or just within lexical choice), Pickering and Garrod (2004) proposed that syntactic and lexical alignment can lead to alignment of the situation model. Evidence exists for the effects of linguistic feedback across different levels of representation: For example, Branigan et al. (2000) found strong priming effects when the prime and target descriptions employed the same verb. Hence, alignment at one level (i.e., lexical) led to greater alignment at another (i.e., syntactic). Cleland and Pickering (2003) found effects of semantic priming on syntactic alignment: Alignment of noun phrase form (e.g., use of a complex noun phrase containing a relative clause, such as ‘the sheep that’s red’) was greater after hearing a noun phrase using the same noun (e.g., ‘the sheep that’s red’) or a semantically related noun (e.g., ‘the goat that’s red’), in comparison to hearing a semantically unrelated noun (e.g., ‘the knife that’s red’). Similar evidence exists for semantic effects on cross-linguistic priming across the languages of bilinguals (Schoonbaert et al., 2007). Thus, it is possible that linguistic information such as labels could affect the coherence of people’s concepts and categories.

However, given that alignment is thought to occur on the basis of priming, it is possible that this priming could occur in the absence of interaction and dialogue and, so, lead to the greater coordination of people’s representations even when those people have not interacted with one another. Specifically, the lexical input of say a certain noun (i.e., as a word label), can increase category coherence across groups of people who have never actually interacted with one another (Malt et al., 1998; Laskowski, 2010).

3.2 Non-interactive alignment

As the literature covered above suggests, alignment is most often discussed as a factor involved specifically in interaction, given its aforementioned use in explaining dialogue processes (specifically, as a process underlying successful dialogue). However, interaction may not always be necessary for alignment, or at least greater coordination between people, since alignment can be considered purely in terms of the matching of information states across individuals, rather than the explicit transfer of information between individuals through interaction (Pickering & Garrod, 2006). For example, individuals may align, or coordinate, without interaction if they simultaneously listen to the same speaker on the radio whilst in different rooms. In this way, both individuals are primed with the same information that similarly affects their representations of a certain notion. This subsequently increases the coherence of representations across the pair, but without each person ever having communicated with the other.

As I discussed in Section 2.3, labels alone can be enough to change people’s categories (Malt et al., 1999; Lupyan, 2008; Laskowski, 2010), supposedly by affecting which dimensions they select for category formation (Waxman & Markow, 1995; Althaus & Mareschal, 2014; Althaus & Plunkett, 2016), and labels can do so in a consistent manner across people (Lupyan et al., 2007; Lupyan & Casasanto, 2014). While most of these studies focused on how individuals categorized items, labels consistently influenced sorters to select certain dimensions for sorting over others, and this might increase category coherence across people. I argue that this in turn could be characterised as a form of alignment, but a form of alignment that does not rely upon our interlocutors interacting – they simply need to receive the same input of, for example, a common label (e.g., an existing word label, a novel non-word label or even a non-linguistic label).

That is not to say, however, that interaction itself is not important in increasing the coherence of people's categories, as clearly it forms a crucial part of how people develop a shared understanding of the world. Given that interaction (e.g., about objects and the categories they belong to) can involve both dialogue generally and more specific aspects like referring labels, how might the effects of dialogue and labels interact with one another to affect category coherence? Markman and Makin (1998) investigated pairs' category coherence for items when they interacted using shared labels for the items, versus when they sorted individually. Specifically, participants had to use plastic building blocks to build a small structure (i.e., a car or a spaceship). Participants in a pair developed shared labels and used these to build the structure collaboratively. In contrast, the other two groups of participants either built the structure alone, or did not build the structure at all. After this building stage, participants then individually categorized the different types of block into groups.

Participants who worked collaboratively with labels showed significantly greater category coherence than those who did not, and crucially their label use was shown to reflect the structure of their subsequent categories (e.g., block shape and colour) in the same manner across both individuals in a pair. As such, the labels (as referring expressions for the blocks) came to act like established names, even after the paired aspect of the task had ended (Brennan & Clark, 1996). It appears, then, that even within dialogue, interaction can help establish common labels, and labels can then bolster coordination between participants to increase category coherence. As such, the factors of both interaction and labelling are important to consider when studying the development of category coherence between people.

3.3 Coordinative versus non-coordinative contexts

Another aspect to consider in terms of labelling and interaction effects is the presence of a partner, or audience, during linguistic categorization. Language is closely tied to communication and to some extent relies on speakers having a shared understanding of how we label objects. The effects of coordinative context and audience design (Clark & Murphy, 1982; Bell, 1984) are subject to the sorter having a partner, or audience, to which they must communicate their categories. Having a partner present during any task and, indeed, interacting with them can thus affect how a person performs that task (i.e., including influences on how they label and, so, linguistically categorize items).

Clark (1985) suggests that interlocutors must identify their common ground or mutual knowledge. This includes the information, beliefs and attitudes that participants know and share (Clark & Carlson, 1982; Clark & Marshall, 1981). As such, some effects of having a partner present on coordination (i.e., between a person and their dialogue partner) relate to modelling the mental state of that person specifically. However, the effects of a coordinative context on coordination between people (specifically here, on category coherence) can also be due to having a task partner in general, that is, not due to the specific partner per se.

How, then, might coordinative contexts affect the way that we use labels to categorize, in way that is not tied to a specific partner? If we are told that our categories should make sense to another person, then the way that we sort (i.e., both with and without labels) must be sensible to others. Within a coordinative context, then, the labels can act as a device for the coordination of categories across people (Clark, 1996). As such, labels (including novel labels) may differentially affect the way that we form categories, such that within a coordinative context labels influence the sorter to select more generalizable and robust category dimensions (Lupyan, 2008; Lupyan & Casasanto, 2014), than if they sorted without labels or if they sorted with labels but in a non-coordinative setting. In this way, labels could increase category coherence across people in coordinative settings, even when sorters do not receive feedback from one another. And what is suitable for one partner, is likely to be suitable across many people, since the need to coordinate with one person could result in the sorter relying upon sorting conventions for that are common for a wider range of people (i.e., beyond the pair; Garrod & Doherty, 1994).

A possible explanation for how such coordinative pressures affect label use is that communication exerts a need for *expressivity* (i.e., the unambiguous relation of a signal, or label, to its intended meaning; Kirby, Cornish & Smith, 2008; Smith, 2003), since we are required to convey a message to a partner, or audience (Pinker & Bloom, 1990; Heine & Kuteva, 2002; Kirby, Tamariz, Cornish & Smith, 2015). This is because in interaction there is a trade-off between word re-use (i.e., pairing one label with one referent to make distinctions between referents clear for the audience), but also the need to make inferential extensions (i.e., extending a label to incorporate new items – a crucial component in categorization). Therefore, label re-use can sometimes contribute to expressivity in a way that does not increase the number of words needed. This is as it allows us to categorize multiple items together under one label, rather than having to use many more labels and, thus, categories (Piantadosi, Tily & Gibson, 2012), and this pressure occurs especially when we have referential communication with a partner, or towards an audience.

Similarly, the presence of a partner or audience can introduce the pressure for optimality: specifically, that word meanings should be optimised for communication. In this way, a person taking into account an audience (be it one partner, or several people), should try to pick labels with the greatest *shareability* (Freyd, 1983). Under the pressure of shareability, we could expect patterns of word use (i.e., as labels) in coordinative settings to be constrained by previously existing conventions which are, thus, more likely to be shared across both the speaker and the audience.

As such, words as labels might only affect categories and category coherence across people in such coordinative situations. Indeed, Silvey (2014) failed to find beneficial effects of labelling upon category coherence between communicating pairs when interactions were computer-mediated and did not involve interacting with a partner directly. Specifically, she examined how novel, non-word labels affected category formation and coherence for novel categories across pairs, with some participants performing the categorization task individually, and others performing it within a collaborative pair (i.e., with the goal of achieving more similar categories to their partner's). She found lower coherence across the categories of collaborative pairs who were negotiating a shared category structure, than between the categories of participants who performed the task individually. Thus, in settings in which the coordinative partner (or audience) is not physically present, it is possible that labels and interaction might fail to increase coherence.

4.1 Bilingual concepts and categories

People's categories can overlap (i.e., be similar), when they use shared features, like perceptual or functional information, as the dimensions for their categories. To a significant extent, language relies on upon *category coherence* across people because people need to have a similar understanding of the world, and how objects within it are labelled, in order to communicate successfully. However, different languages can parse up the world in different ways, leading to different linguistic categories for objects (i.e., differences in how those objects are labelled and, so, in how they are also grouped together under those labels). In addition, bilingual speakers may speak two languages that parse up the world in different ways. In such cases, how do bilinguals resolve the conflict across their languages? Research so far has focused on whether: (a) bilinguals retain separate concepts and categories for their languages (i.e., such that they have similar patterns of categorization to monolinguals in either language), or (b) bilingual linguistic knowledge affects bilinguals' categories and concepts in a way that changes their patterns of categorization in both languages (i.e., such that they differ from monolingual patterns of sorting).

Ameel, Storms, Malt and Sloman (2005) extended the paradigm of Malt et al. (1999) to a bilingual domain, in order to investigate which of the two hypotheses best explained how bilinguals might map their words onto referents across their two languages (see Fig. 5). The two-pattern hypothesis states that bilinguals maintain two distinct mappings from their languages, whereas the one-pattern hypothesis states that bilinguals maintain only one word-referent mapping system for both languages. In the one-pattern hypothesis, this single system of mapping is formed through the merging of the two separate systems of the languages. As such, this one system of mapping diverges from the mapping systems in either of the individual languages, such that a bilingual's word-referent mapping is not native-like in either language, when they are asked to name objects.

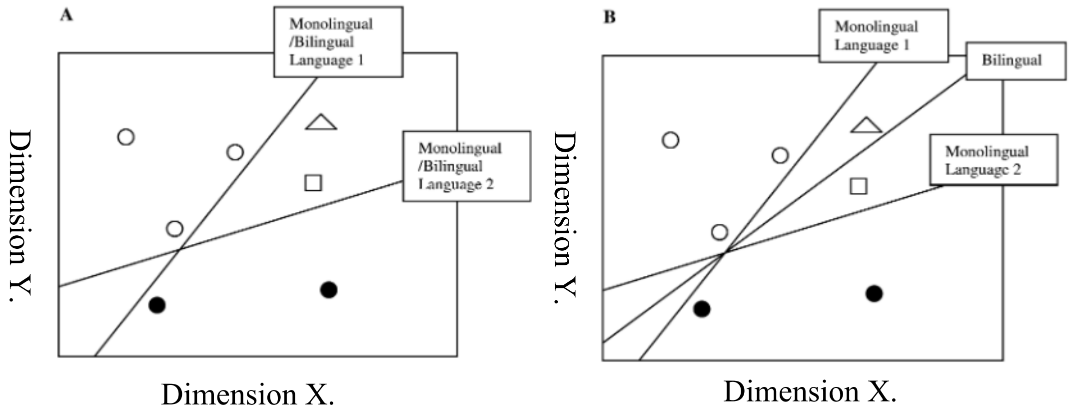


Fig. 5. (A) two-pattern hypothesis and (B) one-pattern hypothesis. Lines represent naming patterns for monolingual vs. bilingual speakers; points represent objects in a semantic space. Adapted from Ameel et al. (2005).

To test these two hypotheses, Ameel et al. (2005) used pairwise similarity judgements (for overall object similarity) and naming data for common, household objects (i.e., containers and dishware). Participants were Dutch or French monolinguals, and Dutch-French bilinguals who had learnt the two languages simultaneously. Also, given that the participants of different language groups were raised in relatively similar cultures, there should have been less cultural differences for the stimuli across the different groups of speakers, than was the case for the different groups of speakers in Malt et al. (1999). Ameel et al. found relatively stable patterns of naming across the monolinguals within each language group, and found significant cross-linguistic differences between the two speaker groups, thus supporting the findings of Malt et al. (1999).

Naming agreement (i.e., group coherence in how speakers labelled the objects) was significantly higher for both the French monolingual and the Dutch monolingual groups, than it was across these two languages (i.e., agreement across Dutch and French monolinguals' naming patterns). However, agreement for the bilinguals was only slightly higher in each of their languages (i.e., bilinguals' agreement within just French, or within just Dutch), than naming agreement across both of their languages. How the bilinguals named objects in Dutch and in French was relatively similar across their two languages, but did not parallel the naming patterns of the Dutch and French monolinguals.

As such, these results were inconsistent with the two-pattern hypothesis that predicted that bilinguals would perform identically to the monolingual groups in both languages. Instead, the results support a moderate version of the one-pattern hypothesis. That is, given that there are not always perfect translation equivalents across Dutch and French, there are some minor deviations in the way that bilinguals name in the two languages. However, overall, there was significant overlap across both languages for bilinguals, and this pattern of overlap was distinct to that of the monolingual naming patterns. And this suggested that bilingual language knowledge had significant effects on the bilingual speakers' concepts and categories.

4.2 L1-L2 interaction and categorization

When two speakers start with significantly different linguistic categories, does interaction still lead to the greater coherence of their categories? So far, I have presented evidence that suggests that between native speakers (L1) of the same language, interacting about and having exposure to another person's categories can aid the development of shared categories and labels. This is because interaction and exposure allow interlocutors to learn how their partner conceives and labels certain concepts, and, thus, how they will categorize objects under a given label (Clark & Brennan, 1991; Markman & Makin, 1998). For example, using the same labels to refer to referents during an interactive, collaborative task can subsequently lead to individuals separately categorizing those referents more similarly to one another (Markman & Makin, 1998).

However, in dialogue between L1 monolingual speakers of the same language there is already a large overlap between L1 speakers, even before we take into account how exposure to and discussion about categories may increase category coherence. The significant differences in the linguistic categories of non-native speakers (L2), compared with L1 speakers, therefore, creates a wider gap to bridge in order to increase category coherence between L1 and L2 interlocutors. This might mean that the development of category coherence through interaction is absent or attenuated in L1-L2 dialogues, compared with cases in which interlocutors have significant category overlap.

Costa, Pickering and Sorace (2008) suggested that non-native representations of linguistic knowledge in L2 speakers may contribute to less coordinated situation models between L1 and L2 speakers, than between two L1 speakers. I posit that this might also be the case for the coherence of linguistic categories before and possibly even after L1-

L2 interaction about categories. Since linguistic categories are concerned with how people label objects, L1 and L2 differences in naming might also contribute to lower category coherence between L1 and L2 speakers of a language (i.e., specifically for linguistic categories, in this case).

Despite these factors, there is evidence that L1s and L2s can coordinate with each other on several levels of linguistic representation, although it is theorized that they may do so through less automatic or implicit routes, than would be expected between two L1 speakers (Costa et al., 2008). For example, L2s could coordinate their choices with the choices (e.g. lexical, syntactic or conceptual) put forth by an L1 speaker in conversation, in order to improve their language acquisition. That is, the L2 speaker recognizes the L1 speaker as a more competent speaker of the language, and therefore attempts to use choices put forth by the L1 speaker. For example, an L2 speaker may switch from calling something a '*cup* of wine', to calling it a '*glass* of wine', after hearing an L1 speaker use the label 'glass'. In this way, it also allows the L2 speaker to then test their usage of a more L1-like choice (i.e., 'glass') against the reaction of the L1 speaker (Mackey, Gass & McDonough, 2000).

L1 speakers could also shift towards L2 speakers for the purpose of accommodation. For example, L1 speakers may try to accommodate L2 speakers by changing their linguistic choices towards simplified speech that is more easily understood by the L2 speaker (Arthur, Wemer, Culmer, Lee & Thomas, 1980). For example, an L1 speaker may call an object a 'cup', instead of a 'mug', when interacting with an L2 speaker, because 'cup' is a more frequent label, and can feasibly apply to mug-like objects (i.e., even if most L1 speakers would prefer the label 'mug' for that particular object).

However, it is also still possible that the simpler speech initially produced by the L2 is implicitly aligned with by the L1 speaker, leading to changes in the L1 speaker's choices (Bortfeld & Brennan, 1997). In this sense, the L1 speaker would call the object a 'cup', instead of a 'mug', simply because they had already heard the L2 speaker use the label 'cup' and, as such, they were primed to use this label. In summary, there are several mechanisms by which L1 and L2 speakers – despite starting with significantly different systems for linguistic categorization – might achieve greater category coherence through exposure to one another's categories and/or interaction about those categories. Additionally, labels may interact with the effects of interaction, to increase the potential for coordination between speakers of different native languages.

5. Roadmap of thesis

Despite perception occurring within the individual, having a similar view of the world - and how to label objects within it - is crucial to successful communication with other people. Multiple factors can affect the structure and similarity of people's categories for objects, by affecting which dimensions they select for forming their categories. Both interaction about categories and using conventionalized labels (i.e., words) when categorizing stimuli can increase people's category coherence (i.e., the similarity of their categories). Despite this, people divide objects into categories in similar ways when asked to do so without language, but when people use language to label groups of objects, significant differences appear across speakers of different languages. What purpose does the act of labelling objects serve, and what factors affect whether labels increase the similarity of people's categories (i.e., categorical coherence)? Do labels change the way we categorize and increase the coherence of our categories alone, or do effects rely on exposure to another person's way of sorting objects using those labels?

In Experiment 1, I investigated whether novel labels increase category coherence, and how this compares to the effects of interaction on category coherence between partners. Pairs of participants repeatedly grouped morphed triangular shapes into two self-determined categories, with or without interim exposure or dialogue to their partner's categories across four conditions. Firstly, pairs in the Non-exposed condition sorted the shapes without labels and without seeing their partner's categories (a). Secondly, for pairs in the three Exposed conditions we manipulated the types of information available to the pairs: (b) pairs sorted with exposure to a partner's categories, but without labels or dialogue, (c) pairs sorted with exposure to a partner's categories and used novel, non-word labels to sort, and (d) pairs sorted with exposure and dialogue, but without the use of labels.

In Experiments 2 and 3, I investigated: (a) whether novel labels also increase category coherence between people, (b) whether novel labels need be linguistic in order to increase this coherence, and (c) whether the effects depended on exposure to a partner's categories. Pairs grouped morphed shapes into two self-determined categories, with or without interim exposure to their partner's categories. I carried out these tasks to investigate the potential effects of labelling and exposure to a partner's categories on coherence, whilst controlling for interaction between sorters. I utilized non-linguistic labels to investigate the potential arbitrariness of labelling effects on categorization.

In Experiments 4 and 5, I investigated whether the effect of labelling on category coherence is dependent on individuals needing to coordinate their categories with other people. This is since labels not only reflect the way that we categorize objects, but also allow us to communicate and share categories with others. In two experiments, I therefore had participants individually sort grayscale images of mountains into two groups with or without novel labels, and with or without a coordinative context.

That is not to say that interaction and exposure to other people's ways of sorting is not sometimes more crucial than the labels themselves. Take, for example, a situation in which two interlocutors do not share the same labels for a group of objects. In these cases when labels cannot be relied on for shared category structure, can interaction with a partner make us more similar to other speakers in a different speech community to our own, native community? In Experiments 6 and 7, I investigated whether interaction in L1-L2 dialogues led to greater category coherence between interlocutors, since L2 speakers' linguistic categories differ from L1 speakers' categories. When two speakers start with significantly different linguistic categories, does interaction still lead to greater coherence of their categories? And if so, what conditions are required to increase this coherence between L1 and L2 speakers? To answer this, I investigated (a) whether discussion increases the similarity of people's categories (i.e., category coherence) in L1-L2 pairs, and (b) how the need for coordination between partners affects this process. L1-L2 pairs individually categorized dishware with intermittent interaction: in Experiment 6, participants discussed categories, or unrelated images; in Experiment 7, all participants discussed categories, but some pairs did so with a coordinative goal.

In my final chapter, I summarize the overall findings of this thesis, as well as the limitations and possible extensions of the methodologies and analyses used. Discussion for Experiments 1-5 will focus mainly on the effects of novel labels and interaction on interlocutors' categories. Experiments 1 and 3 will help us draw some conclusions on the effects of exposure and interaction between people on the coherence of their categories, with and without novel labels. Discussion of Experiments 4 and 5 will add to this, in allowing us to make inferences about how having a coordinative context affects label use in relation to category coherence across sorters. Lastly, discussion of results for Experiments 6 and 7 will switch to focusing mainly on the effects of interaction on linguistic category coherence in L1-L2 dialogues, in order to draw some conclusions on whether interaction can increase category coherence in cases where speakers have very different linguistic categories.

Chapter 2

Use of novel labels increases category coherence more so than discussion about categories

Multiple factors can affect the structure and similarity of people's categories. When discussing categories, both the interaction itself and the use of conventionalized labels (i.e., existing words) can increase people's category coherence (i.e., the similarity of their categories). However, labelling effects may not be limited to conventionalized labels. We investigated whether novel labels increase category coherence, and how this compares to the effects of discussion, between partners. Pairs of participants repeatedly grouped morphed triangular shapes into two self-determined categories, with or without interim exposure to or dialogue about their partner's categories. Pairs without exposure to their partner sorted items without labels and without seeing their partner's categories (a). For pairs who were exposed to their partner's categories, we manipulated the types of information available to pairs of sorters: (b) pairs had exposure but sorted items without labels or dialogue, (c) pairs had exposure and used novel, non-word labels to sort without dialogue, and (d) pairs had both exposure and dialogue, but sorted items without the use of labels. Surprisingly, sorting using novel labels with exposure to a partner's categories led to significantly higher category coherence, than did full dialogue between partners.

Introduction

When people categorize items in the world, they can do so in a number of ways, for example by categorizing on a perceptual, functional or linguistic basis. What strategies people use for categorization, however, can depend on the type of category being formed and the information available to the sorter (e.g., the context of the sorting, interaction between sorters and access to linguistic labels for the items being sorted). These factors can, therefore, also affect the coherence of people's categories. Past research has suggested that discussion about categories can lead to increased coherence between interlocutors, including the increased coherence of people's linguistic categories for items (e.g., Markman & Makin, 1999). There is also evidence that labels alone can change the way that people sort – and have consistent effects upon people's categories – regardless of whether the labels have a conventionalized meaning, by influencing which dimensions people select for sorting (Lupyan, 2008; Lupyan & Casasanto, 2014). However, work that has directly addressed the effect of labels on shared reference use and/or linguistic category coherence across people has mainly examined the use of conventionalized labels within dialogue-based tasks (e.g., Brennan & Clark, 1996; Garrod & Doherty, 1994; Markman & Makin, 1999).

As such, we aimed to compare the effects of using labels when categorizing, with the effects of discussion, on pairs' category coherence. Specifically, we investigated whether encouraging pairs of sorters to use of novel, non-conventionalized labels (i.e., non-word labels that lack a conventionalized meaning within the mental lexicon) when categorizing would affect people's categories in a way that increased the category coherence between people, and how this compared to coherence in cases in which participants were allowed to interact with each other – but without encouragement to use common labels. If novel labels increased coherence only in cases where participants could discuss categories, then our novel labels condition would potentially show lower coherence than the dialogue condition (i.e., since our novel labels condition did not allow for discussion between sorters). However, if the use of novel labels can directly affect the way that people sort – regardless of discussion, and in a way that increases coherence across sorters – then we might see comparable levels of coherence in the novel labels and dialogue conditions.

For comparison, we included a condition in which pairs had interim exposure to their partner's categories, but no labels or dialogue, to address solely the effect of exposure to a partner's categories on category coherence. We predicted that, if labels directly influence categorization in a way that increases coherence and that interaction also improves coherence, pairs in the novel labels and dialogue conditions would show greater category coherence, than pairs in this exposure-only condition. Lastly, we also included a baseline condition in which pairs where never exposed to or interacted about their partner's categories, and in which pairs sorted without labels. If the factors of labelling, interaction and exposure in general were important to establishing greater category coherence between pairs, then all three of these conditions would produce greater category coherence in pairs, than would the non-exposed, baseline condition.

1.1 Labels in Interactive alignment

Previous work has suggested that discussion about categories can lead to people developing more similar categories to each other (i.e., people placing more similar items under the same labels), because exposure to another person's categories and labels affects how a person conceives and labels particular concepts, and so, can lead to people developing more similar categories to each other (Clark & Brennan, 1991; Markman & Makin, 1998). Increased conceptual and categorical coherence through dialogue can be theoretically explained by *Interactive Alignment* (Pickering & Garrod, 2004). Under this account, interlocutors align situation models through dialogue which leads to more similar representations of the topic discussed. This occurs as a product of interlocutors making more similar linguistic choices (e.g., lexical, syntactic or conceptual choices) as they interact with one another. Murphy (2002) argues that categories can be considered as the application of concepts to objects in the real world and, therefore, because interactive alignment leads to increased coherence between interlocutors at the conceptual level, it is possible that dialogue may also lead to people having more similar categories of items (e.g., in terms of how interlocutors would label and, therefore, linguistically categorize items; Clark & Brennan, 1991).

Markman and Makin (1998) had pairs of participants perform a joint task in which they had to construct objects from building blocks by following given instructions. Pairs were either encouraged to use common labels to refer to the different types of building blocks, or not encouraged to use common labels, during this process. Participants in the control condition built the objects individually, without labels. Following the building task, participants then had to individually categorize the types of building block into different groups, and these groups were used to measure category coherence between pairs of participants. Participants who had used common labels to refer to the building blocks during the building part of the task had more similar categories to their partner, compared with those who were not encouraged to use common labels, and compared to participants who built the objects individually. These results suggest two explanations that might lead to the increased coherence of people's linguistic categories: a) interaction increases coherence over cases in which people do not interact; and b) using common labels to refer to categories increases coherence, compared with not using common labels. However, from these results it is also possible that labels only increase coherence in cases where people have some form of interaction with one another.

1.2 Labels in non-interactive alignment

Other research has addressed the potential effects of labelling on category coherence in the absence of interaction, and it appears that labels can indeed affect people's categories in consistent ways. People categorize differently when they use word labels to sort items (i.e., linguistic categorization), compared with when they sort without labels. And linguistic categories vary across speakers of different languages. For example, when categorizing the same set of stimuli (i.e., 60 container-like objects), a group of native Chinese Mandarin speakers tended to use five different word labels (i.e., corresponding to five linguistic categories for the items), whereas American English speakers used seven and Argentinian Spanish speakers used 15 (Malt et al., 1999). Conventionalized word labels also have an important impact on category induction: Older children and adults rely more on common labels than perceptual coherence when deciding an object's category membership (Sloutsky, Lo & Fisher, 2001). People can therefore sometimes categorize objects in the same way as each other because they use a common set of linguistic labels.

Evidence from developmental studies has affirmed the importance of linguistic labels on category learning in early childhood, and may help explain the mechanisms underlying the effects of labelling on categories in adults. Althaus and Mareschal (2014) examined how labels might facilitate category learning in children, and whether this effect was contingent on linguistic input. Eye-tracking was used to examine toddlers' fixations to certain features across novel, alien-like stimuli. Results demonstrated that, when stimuli were presented alongside novel labelling phrases (e.g., 'Look at the Timbo!), infants more rapidly focused their attention to common category dimensions across the stimuli, than when stimuli were presented alongside non-labelling phrases (e.g., 'Look at that!'). These findings are also consistent with evidence that labels can affect early visual processing in adults (Lupyan, 2008; Lupyan & Spivey, 2010; Boutonnet & Lupyan, 2015).

Lupyan (2008) suggested that labels can also affect categorization in adults, by altering existing representations for objects and, therefore, distorting memory for the physical features of an object most reliably associated with a category label. This suggests that labelling affects representations in ways that might also affect categorization. In a recall test, adult participants were worse at recognizing items of furniture that they had previously seen if they had named the object (e.g., they identified it as a 'table' or 'chair') than if they had made a like/dislike decision for it. They performed most poorly when recognizing previously labelled objects that were highly prototypical and unambiguous (i.e., the objects with the fewest idiosyncratic features to differentiate them from common features associated with a category). Lupyan proposed that when speakers apply a category label to an object, it causes the representation of the item to become a mix of its idiosyncratic features and the features typically associated with the category. This makes it harder to recall the object accurately, a tendency that is exacerbated when the object has few discerning features.

Representational shift may lead to lower accuracy when recalling specific objects, but as Lupyan (2008) noted, it helps people to select category dimensions that can be successfully abstracted across many objects within a category. This in turn helps people to avoid forming categories on too specific or too fine a set of dimensions, which would lead to a greater and potentially unhelpful number of categories. For example, good factors for identifying whether an item should be categorized as a chair could be something with (generally) four legs and something that you can sit on. Poor factors for making this decision would be the colour of the material it's made from, or whether it is cushioned.

For people within a language group, labels seem to increase the likelihood of them selecting more similar category factors (Laskowski, 2010). This shift acts a re-organization of what category features the sorter prioritizes when deciding how to divide up the items. The relevant question here is whether this makes people tend to reorganize in predictable ways, and ways that tend to be similar across people.

In summary, if labels directly affect how people categorize, then labels might also increase category coherence between people without the need for the explicit transfer of information between sorters (e.g., without the need for interaction). Under this non-interactive account, labels might lead to increased coherence between people even in the absence of interaction or dialogue, by providing a common input to each person individually that can consistently affect different people's selections of category dimensions.

1.3 Non-conventionalized, novel labels

It is clear that the use of existing, conventional labels influences how people categorize objects. But people frequently encounter new labels for novel or established objects. For example, names for new technology can require the learning of new words (e.g. 'dongle'), whilst second language acquisition (e.g., acquisition of nouns) requires the learning of new words as names for existing objects. Are categorization processes in adults similarly influenced by the use of such novel, non-conventional labels, or must labels be conventionalized in order to influence categorization?

Lupyan and Casasanto (2014) had participants categorize novel alien-like stimuli into two pre-determined categories using conventional labels (i.e., 'smooth headed' or 'pointy headed') or non-word labels that involved some degree of sound symbolism (i.e., 'fooves' or 'crelches'; cf. Köhler, 1929). Participants performed equally well in learning to assign stimuli to the relevant categories whether they used conventional or novel, non-word labels. Performance for both label types was better than in the control condition, in which participants sorted items without labels. These results suggest that labels can facilitate category learning in ways that do not depend upon access to conventionalized form-meaning pairings in the mental lexicon (see Jackendoff, 2002). Instead, it appears that the use of novel labels drew focus to the necessary perceptual features (i.e.

pointedness or smoothness cues) that supported category learning within individual participants.

Another way that novel labels might affect category formation and lead to increased category coherence across people is by simplifying the reference to item features that form the category dimensions. Lupyan et al. (2007) investigated the efficacy of printed and spoken non-word labels (i.e., leebish or grecious) compared with other non-linguistic cues in categorizing novel stimuli (i.e., participants categorizing aliens as 'friendly' or 'unfriendly'). Both printed and spoken word labels facilitated category learning. Lupyan et al. (2007) argued that word labels were effective because they simplified the distinction between categories. That is, using word labels meant that participants could categorize the objects under a single term (here, leebish or grecious) that represented the necessary category dimensions (e.g., 'spiky' and 'with ridges') and made these dimensions more concrete. In contrast, they propose that when people did not use labels, they had to rely on more complex and fuzzy perceptual distinctions (e.g., objects categorized as 'more rounded and smooth' or 'less rounded, with ridges').

Similarly, Edmiston and Lupyan (2015) argued that linguistic labels are advantageous for processing category members because they are superordinate. That is, words can be abstracted over all kinds of items that fall under that label, despite some disparities in features across items (e.g., whilst some dog breeds are considered very large and other breeds are comparatively very small, all dogs can still be referred to using the label 'dog'). In comparison, using environmental sounds that give more specific information about an item (e.g., a high-pitched yap) plausibly refer only to small breeds of dog. Thus, labels, even novel ones, may be unique in allowing the sorter to more easily abstract category dimensions over a wider range of items.

Lastly, in previous research examining the effects of labelling on categories – both with and without interaction – participants received repeated exposure to the items they were categorizing and the labels they used for categorization (e.g., Markman & Makin, 1998; Lupyan & Casasanto, 2014). As such, the repeated exposure to both a partner's categories and their labels might be necessary for increased coherence between sorters, since it allows more time for pairs to learn how a partner conceives and labels particular concepts, and hence to develop categories that are similar to theirs. If this is the case, there might also be an effect of round (as well as type of exposure between partners) on the coherence of a pair's categories (i.e.., increased category coherence across rounds in exposed conditions). If not, then the effect of labels could be something more immediate, that is, having labels present during category formation impacts categorization to create a greater state of coherence. In this scenario, repeated exposure to a partner's categories and labels would not be necessary for the labels to cause increased category coherence. Instead we would expect to see an overall increased state of coherence in general when people sort with labels than when they sort without labels.

1.4 Current study

Research has shown that exposure to and discussion about categories can increase category coherence between sorters. However, when people categorize objects, they are also affected by linguistic factors. Both conventionalized and novel (i.e., non-word) labels can have consistent effects on the way that people categorize (i.e., by influencing which dimensions people select for sorting). Labels do so potentially by drawing attention to more consistent and generalizable features in a manner that can lead to coherence across people's categories. But we do not know whether novel, non-conventionalized labels can affect self-determined categorization across people in a consistent manner, and whether exposure and dialogue between people is necessary for novel labels to affect category coherence in this way.

In this study, pairs of participants sorted identical sets of geometric stimuli (i.e., morphed, monochromatic triangles that varied in size, shape, angles, and pointedness) into two self-determined categories over six rounds. Firstly, we manipulated whether participants were exposed to their partner's categories or not. Pairs without exposure to their partner sorted items without labels and without seeing their partner's categories (a). For pairs who were exposed to their partner's categories, we manipulated the types of information available to pairs of sorters: (b) pairs had exposure but sorted items without labels or dialogue, (c) pairs had exposure and used novel, non-word labels to sort without dialogue, and (d) pairs had both exposure and dialogue, but sorted items without the use of labels.

We predicted two possibilities for the effects of novel labels on category coherence across pairs of sorters: (a) if the effects of novel labels were dependent on discussion between interlocutors, then pairs in the Exposed-with-labels condition would not have greater coherence than in the Exposed-no-labels, or Non-exposed conditions, and Exposed-with-labels might produce lower category coherence than the Exposed-with-dialogue condition, but (b) if novel labels can directly influence both sorters in a pair to select more similar category dimensions to one another, then coherence would be greater in the Exposed-with-labels condition, than in the Exposed-no-labels and Non-exposed conditions, and would be at least comparable to coherence in the Exposed-with-dialogue condition.

Methods

2.1 Participants

Participants were 64 native, monolingual British or North American speakers of English (49 female), who were assigned to 32 pairs. Participants within a pair did not know each other. Participant ages ranged from 18-30 years (M = 22.34, SD = 2.72). The University of Edinburgh's Ethics Committee approved this study.

2.2 Stimuli

Stimuli were six sets of 24 monochromatic triangular shapes (henceforth, triangles) which varied in the length of sides (and therefore angles) and in the degree of pointedness of vertices (see Fig. 1; see Appendices A-F for the full sets). Triangles were printed on 3x3 inch cards. Additionally, we created six pairs of non-word labels (e.g., 'WEF' or 'GIS'; see Appendix G for the full list of non-word labels) for use in the Exposed-with-labels condition. Labels remained the same for participants across rounds.

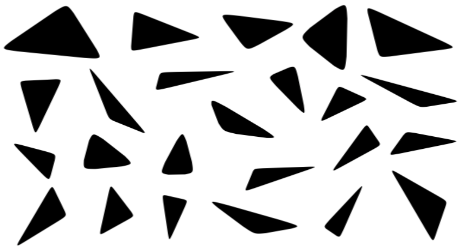


Fig. 1. Example set of morphed triangular stimuli.

2.3 Design

There were two independent factors: Exposure (between-participants; Non-exposed, Exposed, Exposed-with-labels, and Exposed-with-dialogue) and Round (within-participants; rounds 1-6). In the Exposed, Exposed-with-labels and Exposed-with-dialogue conditions participants viewed their partner's categories after each round of sorting. In the Non-exposed condition, participants were never exposed to each other or their categories. Six pairs were randomly assigned to each of the four exposure conditions.

2.4 Procedure

Participants were seated opposite one another with the barrier in place (see Fig. 2), and sorted one set of stimuli per round into two categories, with a time limit of 3 minutes per round. All triangles had to be assigned to a category, and each category had to contain a minimum of 6 and a maximum of 18 triangles. Our reasoning was that requiring a specific number of triangles per category (e.g., 12 each) could have led to participants 'forcing' triangles into categories that they deemed inappropriate, and allowing complete freedom could have led to very small categories. Lastly, participants were informed that they were free to sort using the same criteria across rounds, or to change the criteria across rounds as they wished.

For participants in the Exposed-with-labels condition, printed labels were presented alongside the stimuli cards. Upon receiving the labels participants were told 'You have these two labels to place upon your categories. Place one label on one category each. You choose how to use them. You can move them across the categories between rounds if you wish'. For all conditions, there was a 1 minute interval between sorting sets. During this interval, the barrier was removed for participants in the Exposed, Exposed-with-labels and Exposed-with-dialogue conditions so that they saw their partner's categories (and labels if they were in the Exposed-with-labels condition); participants in the Non-exposed conditions did not see their partner's categories. Instead, they were asked to silently reflect upon how they had categorized during this interval. Talking between participants was allowed only in the Exposed-with-dialogue condition, and only during intervals.

Following the interval, the barrier was then replaced for Exposed conditions, and was never removed in the Non-exposed conditions. The sorted cards were collected for recording results and a new set of stimuli cards was provided. The order of presentation of sets was counterbalanced across pairs. Participants completed a post-test questionnaire about the task at the end of the experiment.

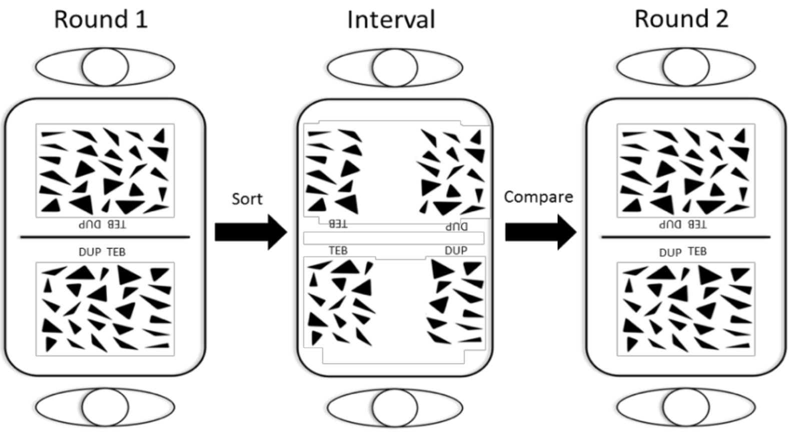


Fig. 2. Example of two rounds of sorting with labels and an interval in the middle (Exposed-with-labels condition).

Results

3.1 Calculating category coherence scores

Scoring reflected how similarly two participants in a pair split their items into two categories (see Fig. 3 for an example of scoring). We arbitrarily called the two categories for one participant A1 and A2, and for the other participant B1 and B2. We counted (a) the number of items that fell into both A1 and B1, and added these to the number of items that fell into both A2 and B2 (i.e., items that both participants placed within their respective categories); we then counted (b) the number of items that fell into both A1 and B2, and added these to the number of items that fell into both A2 and B1 (see Fig. 4). The category coherence score given to a pair was the larger out of (a) and (b). This meant that the minimum score was 12 and the maximum score was 24.

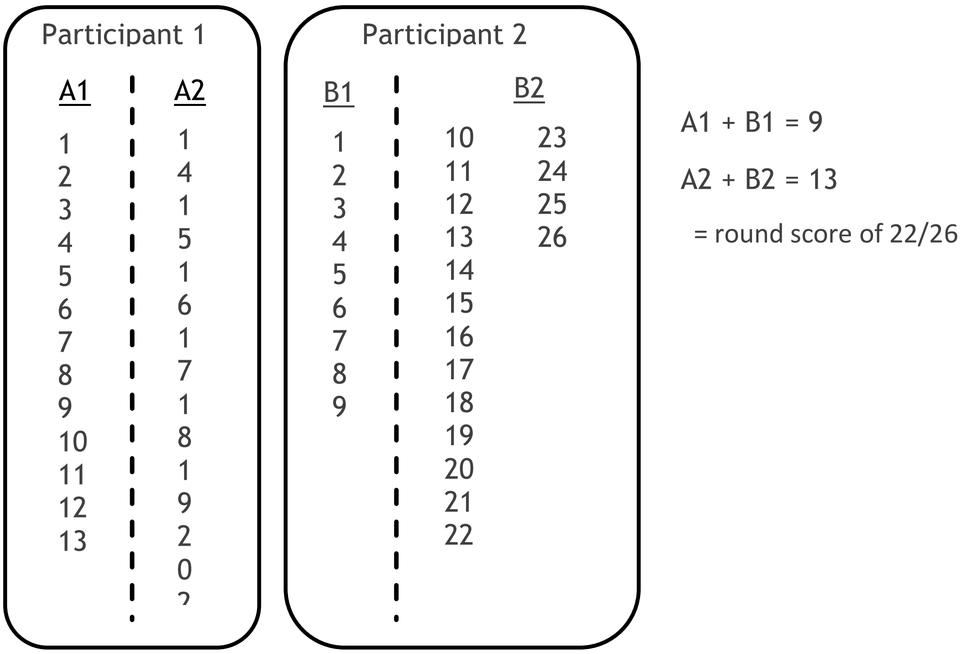


Fig. 3. Example of scoring in which this pair of participants would score 22/26 for category coherence in this round. Each number refers to items 1-26.

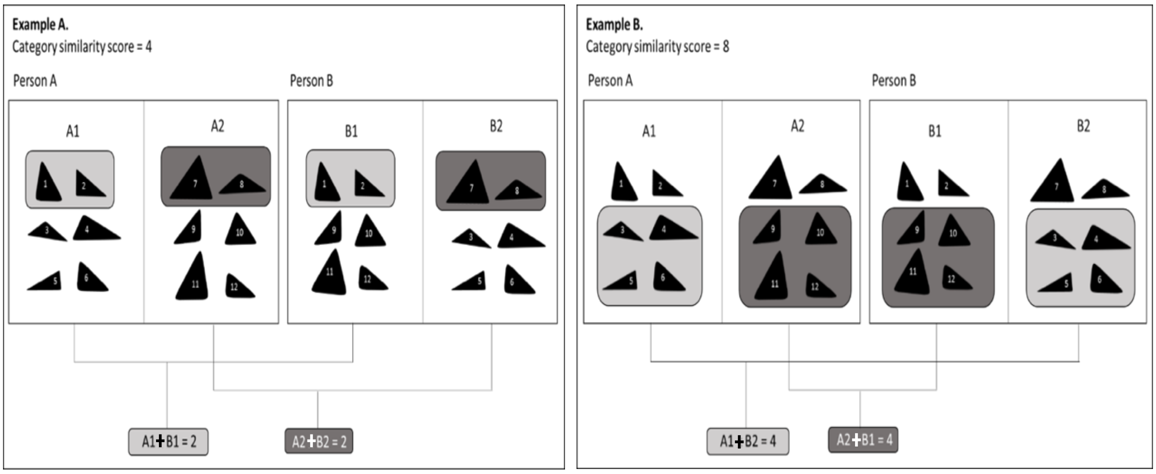


Fig. 4. Examples of category coherence scoring. Example A reflects minimal scoring, while B reflects maximal scoring. Maximal scoring was used to calculate all scores.

3.2 Descriptive statistics

Table 1 reports the mean category coherence scores across pairs for the conditions Exposure by Round (1-6), and the overall grand mean for each condition of Exposure (i.e., averaged across all six rounds).

Table 1.

Mean coherence scores and grand mean (SD) for Exposure by Round (1-6).

	Non-exposed	Exposed-no-labels	Exposed-with-labels	Exposed-with-dialogue
Round 1	17.33 (3.01)	16.50 (1.97)	20.17 (2.56)	16.33 (3.56)
Round 2	18.17 (3.41)	18.17 (4.79)	18.83 (3.54)	16.17 (3.31)
Round 3	16.67 (3.44)	17.67 (3.93)	20.17 (2.04)	16.67 (2.08)
Round 4	18.00 (1.79)	18.83 (2.99)	19.17 (2.64)	16.00 (2.53)
Round 5	16.00 (1.79)	17.83 (3.31)	19.67 (2.16)	16.83 (2.86)
Round 6	17.82 (2.04)	19.00 (1.55)	20.50 (1.54)	17.00 (3.29)
Grand $\bar{x}$	17.33 (2.58)	18.00 (3.09)	19.75 (2.41)	16.50 (2.94)

3.3 Main analysis

A mixed-design analysis of variance (ANOVA) was computed using SPSS 21 (IBM Corp., 2012). The factor of Exposure (Non-exposed, Exposed, Exposed-with-labels, and Exposed-with-dialogue) was entered as a between-subjects factor, whilst Round (referring to coherence scores at each interval between rounds 1-6) was entered as a within-subjects factor. The interaction between Round and Exposure was not significant ($p > .05$). There was also no main effect of Round on coherence scores ($p > .05$). However, there was a significant main effect of Exposure on coherence scores ($F(3,28)=4.60$, $p < .05$). Thus, the final model included only Exposure as a predictor (see Table 2). A post-hoc Tukey test revealed that the effect of Exposure was due to category coherence scores for the Exposed-with-labels (M = 19.75, SD = 2.41) being significantly higher than category coherence scores for Exposed-with-dialogue (M = 16.50, SD = 2.94) ($p < .01$).

Table 2.

ANOVA results for effects of Exposure on category coherence.

	df	SS	MS	F	p
Exposure	3	233.50	77.83	4.60	.01
Error	28	474.08	16.93		

3.4 Labels-specific analysis

We next examined whether the way that participants assigned each of the labels to the triangles mattered for category coherence, in specifically the Exposed-with-labels condition. This done to give us an indication of whether the labels were being applied to certain stimuli in a non-arbitrary way. Coherence scores for each round were assigned to a ‘matched’ or ‘non-matched’ condition dependent on whether participant A placed specific items under ‘WEF’ and others under ‘GIS’ in a similar fashion to how participant B did. There were 24 cases in both the ‘matched’ (M = 19.63, SD = 2.65) and ‘non-matched’ (M = 19.17, SD = 2.73) conditions. An independent samples t-test was used to assess whether there was a significant difference in category coherence scores due to the novel labels being matched or unmatched across rounds. Results demonstrated the difference to be non-significant ($p > .05$).

Discussion

Across six rounds, pairs of participants repeatedly grouped morphed triangular shapes into two self-determined categories, with or without interim exposure to their partner's categories. Participants in the Non-exposed condition did not discuss categories and were never exposed to their partner's categories during the intervals between rounds. Participants in the Exposed condition were in either: Exposed-no-labels – pairs did not discuss categories with their partner, but were allowed to see their partner's categories at intervals (i.e., during the intervals between rounds); Exposed-with-labels – pairs did not discuss categories with their partner, but sorted the shapes using novel labels and saw their partner's categories and labels during the intervals; or Exposed-with-dialogue – pairs sorted the shapes without labels, but saw their partner's categories and were able to engage in full dialogue about their sorting strategies with their partner at the intervals.

We predicted that if the use of novel labels directly affected the way that people sorted – regardless of discussion, and in a way that increased coherence across sorters – then we might see comparable levels of coherence in both the Exposed-with-labels and Exposed-with-dialogue conditions. Secondarily, we predicted that, if labels did directly influence categorization in a way that increased coherence, then pairs in both the Exposed-with-labels and Exposed-with-dialogue conditions would show greater category coherence, than pairs in the Exposed-no-labels condition. However, if novel labels increased coherence only in cases where participants could discuss categories, then the Exposed-with-labels condition would potentially show lower coherence than the Exposed-with-dialogue condition (i.e., since the Exposed-with-labels condition did not allow for discussion between sorters). Lastly, we predicted that – if the factors of labelling, interaction and exposure in general were important to establishing greater category coherence between pairs – then all three of these conditions would produce greater category coherence in pairs, than would the Non-exposed, baseline condition.

Results demonstrated that pairs of participants who sorted with novel labels (Exposed-with-labels) had greater category coherence than pairs who sorted without labels but with full dialogue (Exposed-with-dialogue). We discuss two possible interpretations of these results: either dialogue reduced category coherence between people, or – more in line with our original predictions that novel labels can directly influence how people select dimensions for sorting – non-conventionalized, novel labels affected how people formed self-determined categories in a way that led to greater category coherence for pairs of people (i.e., compared to the pairs who had dialogue). There were no significant differences in category coherence between any of the other conditions of Exposure, and no effect of Round over time. Lastly, within specifically the Exposure-with-labels condition, there was no significant difference in coherence scores whether two participants in a pair assigned the same novel label to similar triangles, or not.

4.1 Dialogue hindered category coherence

Given that dialogue between partners allows not only exposure to a partner's categories, but also explicit discussion of each person's criteria for sorting – we predicted that pairs who discussed their categories would have the greatest category coherence of any of the conditions of Exposure. However, this was not the case, with pairs who had dialogue having significantly less category coherence than pairs who sorted with shared novel labels and no dialogue. Markman and Makin (1998) found greater category coherence between pairs that had interacted using common labels, than those who had interacted without the use of these common referring terms. They argued that these common labels helped sorters select object features that worked well across a range of items, and helped interlocutors to coordinate their categories. And within a coordinative context, such as a paired task, labels can act as a device for the coordination of categories across people (Clark, 1996). Perhaps then, the issue with our Exposed-with-dialogue condition is that the dialogue prevented participants setting up labels, such that they did not develop a common system of labels (i.e., especially since the use of common labels was not explicitly promoted between participants). This effect might also have been compounded by the constraints on the amount of interaction pairs had between rounds. And on the other hand, within such a short task simply giving participants pre-defined labels (i.e., even novel, non-words) may have been more beneficial to the development of more similar categories, than dialogue without pre-defined labels.

Another possibility is that audience design (i.e., here, trying to adopt sorting strategies that would make sense to someone else) might have influenced individuals to use less preferred strategies, than if they need not have taken another person's perspective into account. For example, Silvey (2014) examined category coherence between pairs who were allowed different levels of computer-mediated interaction and found lower coherence scores between communicating pairs who were negotiating category structure, than between non-communicating pairs (i.e., pairs of participants who had individually sorted the items). Although Silvey's communicative condition did not allow full dialogue, they argued that having access to a partner's category system, as well as the individual's own system, could have introduced a greater cognitive load (i.e., through each person in a pair having to store and update two systems simultaneously). They added that this in turn could have introduced noise into the categorization process, which hindered the development of greater category coherence between people. However, the communicative condition of Silvey's research can be considered closer to that of our exposed without dialogue conditions, thus, making the comparison with our results less straightforward. As Clark and Brennan (1991) argued, co-presence (i.e., the presence of both interlocutors in the same space) is a crucial feature of face-to-face interaction, and, so, might be considered integral to coordination between people. This co-presence was absent in Silvey's study (i.e., since communication was computer-mediated) and, so, may explain the limitations of the communicative condition in supporting the development of coherence between partners. While participants in our study did have co-presence, exposure between them was limited to short, restricted intervals between sorting.

Lastly, it is possible that – without the explicit goal of coordinating their categories – pairs of participants did not explicitly discuss and negotiate what strategy of sorting they would use in the coming round, following each interval. As such, they could have repeatedly swapped between each other's strategies from round to round, whilst never managing to coordinate on both using one strategy within the same round. A similar issue could be that using novel stimuli for each round made it too difficult to apply strategies across rounds. So, when participants did use dialogue to try and adopt the same strategy as each other in the next round, they failed to apply this strategy to the new stimuli in the same way as each other. In the future, using the same items across rounds could reduce task difficulty and allow participants to more successfully apply shared strategies to items across multiple rounds.

4.2 Labels increased category coherence

In addition to an account in which dialogue hindered the development of more similar categories, our results could also be interpreted as evidence that novel labels can have consistent effects upon people's categories, even in the absence of interaction between sorters, potentially by influencing sorters to select less individualistic, more perceptually-common dimensions for sorting (Althaus & Mareschal, 2014; Althaus & Plunkett, 2016; Lupyan et al., 2007; Lupyan & Casasanto, 2014). A qualitative example of shared versus individualistic strategies from Experiment 1 is the contrast in wide-spread strategies reported by many participants and across several conditions (e.g., 'BIG' vs. 'SMALL'), and individualistic strategies such as Participant X's (dialogue condition) strategy of 'Triangles you'd find on a beach' versus 'Triangles you wouldn't find on a beach'. Whilst this participant's strategy still uses reference to perceptual features to some extent (i.e., one factor in her decision-making was the sharpness of the triangle corners), there are several aspects of this strategy that make it individualistic (such as the context) and, therefore, unlikely to be a strategy employed by other sorters. In this sense, the room for negotiation in our dialogue condition may have led to the choice of more individualistic strategies (Silvey, 2014), whilst the lack of dialogue but the addition of labels pushed sorters towards more shared, perceptual strategies in the labelling condition (Lupyan, 2008). Moreover, our results would support the notion that increased category coherence with novel labels does not depend on repeated exposure to a partner's categories, or extended experience of exemplars across rounds (i.e., because the effect of labels on coherence did not increase across rounds).

As such, we conducted a labels-specific analysis to further examine the effects of novel labels within the Exposed-with-labels condition. There was no significant difference in category coherence whether two participants in a pair assigned the same novel label to similar triangles, or not. Therefore, the content (i.e., any meaning given to the labels by participants) did not appear to be driving the difference in scores between the Exposed-with-labels and Exposed-with-dialogue conditions. Given that content did not appear to be driving the labelling effects, an alternative explanation for the greater coherence of pairs sorting with novel labels, than with dialogue, is that labelling directs people's attention to common features across objects. This might occur through labels supporting the selection of more abstractable, perceptual category dimensions in a consistent manner across people (Waxman & Markow, 1995; Lupyan, 2008). Just as there are similarities in the way people perceptually experience objects, there are fundamental commonalities in which dimensions people find easy to abstract (Rosch & Mervis, 1975; Johnson, 1987). Hence, even novel labels could potentially support the selection of more abstractable, perceptual category dimensions, and this selection is likely to be common across people. In other words, labels could lead to greater category coherence, because they help people to identify appropriate category dimensions – and what is appropriate for one person is likely to be appropriate for any other person.

Lastly, novel labels might have influenced people's categorization by simplifying the reference to category distinctions, as posited by Lupyan et al. (2007). For example, a novel label such as 'GIS' could have come to represent the combination of two distinct dimensions (e.g., 'thin' and 'pointed' triangles). In doing so, it may have strengthened this combination of dimensions (i.e., made it more concrete) and so made it more easy to remember and apply to new exemplars, than if participants had to remember these category dimensions separately and without a common label to represent the two dimensions (Lupyan et al., 2007). Thus, while we cannot say whether dialogue hindered category coherence or whether labels increased category coherence, there seems to us to be a substantial basis for the importance of labelling effects on categorization and coherence, even without interaction between sorters.

4.3 Category coherence across rounds

Lastly, there was no effect of round on coherence (i.e., there was no significant trend in the coherence of pairs' categories moving from round one to six). This lack of increase or decrease of coherence across rounds provides no evidence that repeated exposure across rounds is necessary for partners to converge on each other's ways of categorizing and labelling. Therefore, our results do not appear to support accounts that have suggested that repeated exposure to a person's categories leads to the emergence of more similar linguistic categories across people over time (e.g., Markman & Makin, 1998; Garrod & Doherty, 1994).

Instead – and although exposure over time may sometimes increase category coherence – the labelling effects found in our study can be considered more immediate, or direct. That is, the act of labelling categories led to an increased state of category coherence between partners, rather than a process of increasing coherence that occurred across rounds and over time. Thus, our results support an account in which labels increase category coherence without the need for information to be repeatedly transferred between interlocutors. The results, therefore, also provide support for an account in which novel labels increase category coherence across people by directly affecting what dimensions they select and abstract when forming their categories, even in the absence of interaction between people (Waxman & Markow, 1995; Lupyan, 2008).

4.4 Conclusion

Past research has shown that labels are more than simply arbitrary words assigned to items with no further effect; they can affect the way items are represented and categorized. Pairs of participants produced more similar categories when they did not communicate and used novel labels to sort, than when they communicated but did not use novel labels to sort. How participants in a pair applied each of the labels to specific items did not affect how similar a pair's categories became, suggesting a direct effect of labelling, which influenced people to select more similar category dimensions to each other. There was also no increase or decrease of category coherence over time, again suggesting that the labels produced a state of greater coherence, rather than through a process by which labels gradually increased coherence across rounds. To better understand the effects of novel labels on category coherence across people, future research should focus on category coherence specifically under conditions with and without novel labels, whilst controlling for interaction between sorters.

Chapter 3

Effects of novel linguistic versus non-linguistic labels and exposure on people's category coherence

Exposure to partners' categories can be crucial to *category coherence* (i.e., coordination of people's categories). However, using conventionalized labels (i.e., words) to categorize stimuli can also increase category coherence. Additionally, novel labels have been shown to have consistent effects upon people's categories. We investigated whether: (a) novel labels increase category coherence between people, (b) novel labels need be linguistic to increase coherence, and (c) effects depend on exposure. Pairs grouped morphed shapes into self-determined categories, with or without interim exposure to a partner's categories. Firstly, we examined coherence between pairs of participants who performed the task together (*pair coherence*). Secondly, we examined coherence between all sorters within a condition (*group coherence*). Despite the results of Experiment 1 suggesting that there may be a general, arbitrary effect of labelling on categorization, there was no significant effect of labels in any of the analyses across Experiments 2 or 3. Instead, there was a significant effect of having exposure to a partner's categories for both pair and group coherence in Experiment 2, and for group coherence in Experiment 3. These results suggest that: a) novel labels do not always benefit category coherence between sorters, and b) that exposure to another sorter's categories may increase coherence between not only the pairs of individuals that were exposed to each other's categories, but that exposure to another sorter may produce a more average way of sorting – thus, increasing the sorters similarity to other sorters to whom they were not exposed.

Introduction

When we perceive the world, we can carve it up in different ways. Why do we sometimes do so in a similar way to other people, and sometimes not? Imagine pairs of people who sort a series of geometric shapes into categories: There may be similarities in their categories, but the exact criteria that they each select for sorting (e.g., a combination of size, length of sides or angles) can vary. One way that people's categories can become more similar is through exposure; that is, being exposed to each other's ways of categorizing objects (e.g., Markman & Makin, 1998). Assigning labels to categories can also increase similarities in the way people sort objects. Existing, conventionalized word labels essentially specify what rules the sorters should use to categorize (e.g., 'big' vs. 'small' as labels). However, work has shown that novel labels (e.g., non-words) that lack conventionalized meanings can also affect people's categories in consistent ways (Lupyan & Casasanto, 2014), which suggests that the act of labelling itself may influence how people categorize.

We investigated whether two forms of novel label (i.e., linguistic non-words and non-linguistic coloured tags) increase category coherence, and whether effects were dependent on exposure to a partner's categories. Pairs of participants grouped morphed shapes into two self-determined categories, with or without interim exposure to their partner's categories. We utilized two measures of category coherence: Firstly, we examined category coherence for pairs who performed the task together (*pair coherence*) – both with and without exposure to a partner's categories – in order to allow us to assess the effects of labelling when sorters are exposed to each other's category systems, or not; Secondly, we examined category coherence across all participants within an experimental condition (*group coherence*) to assess whether there were generalizable effects of labelling and exposure on category coherence across people, in addition to between the experimental pairs. This group coherence therefore reflected whether sorters became more similar to an average way of sorting as a group (i.e., on the basis of having labels and having exposure to a partner, or not).

1.1 Factors affecting categorization

Several factors can affect how people categorize objects, and lead to coherence in categories across people. Objects' perceptual features (i.e., their size, shape, and colour) are crucial to category formation in pre-linguistic infants (Murphy, 2002), and adults show remarkable agreement in how they categorize objects using perceptual features, even when they come from different language backgrounds (Malt et al., 1999; Laskowski, 2010). Thus, people may sometimes categorize objects in the same way as each other because they share perceptual experience of the same, structured world (Rosch & Mervis, 1975; Johnson, 1987).

Non-perceptual features like object function can also play a critical role in categorization. Object functions can draw infants' attention to perceptible features and, thus, aid categorization of objects that share such features by around 14-months of age (Waxman & Booth, 2001). Infants as young as 18-months are able to correctly select category members on the basis of shared, salient function (Booth, 2006), and function continues to affect categorization in adulthood. Thus, as with perceptual experience, people may categorize objects in the same way as each other because they share functional experience of those objects.

But categorization is also affected by language: People categorize differently when they use word labels to sort items (i.e., *linguistic categorization*), compared with when they sort without labels. And linguistic categories vary across speakers of different languages. For example, when categorizing the same set of stimuli (i.e., 60 container-like objects), a group of native Mandarin speakers tended to use five different word labels, whereas English speakers used seven and Spanish speakers used 15 (Malt et al., 1999). Conventionalized word labels also have an important impact on category induction: Older children and adults rely more on shared labels than shared perceptual features, when deciding objects' category membership (Sloutsky et al., 2001). People can therefore sometimes categorize objects in the same way as each other because they use a shared set of linguistic labels for those objects.

1.2 Effects of conventionalized labels on categorization

One explanation for the effects of conventionalized labels on categorization is that they drive top-down processing of object properties. For instance, hearing that an object is a ‘turtle’ drives us to attribute to that object properties that we associate with other, previously encountered objects that have also been labelled as ‘turtle’, properties such as being reptilian and cold-blooded (Gelman, 2003; Gelman & Davidson, 2013). Such effects extend to influencing early visual processing during categorization. For example, Lupyan and Spivey (2010) showed that hearing a redundant label (i.e., a label that provides no new information relevant to the task) still effectively guides visual attention towards target objects. During a visual identification task, people were faster to locate 5s (in a display of 2s & 5s) after hearing ‘five’ than was the case on trials in which participants did not hear this redundant label.

Secondly, Lupyan (2008) suggested that labelling affects categorization by causing a shift in how people represent categories, and that labelling does so by distorting their memory for the physical features of an object most reliably associated with a category label. That is, when speakers apply a category label to an object, it causes the representation of the object to become a mix of its idiosyncratic features and features typically associated with the relevant category. Consistent with this claim, participants were worse in a recall task at recognizing objects of furniture that they had previously seen if they had labelled the object (e.g., they had identified it as a ‘table’ or 'chair') than if they had made a like/dislike decision for it.

Lupyan argued that this representational shift supports categorization, by helping people to select more abstractable category dimensions (i.e., consistent and generalizable dimensions that work well across a range of objects within the category). They are thus able to avoid forming categories on too numerous and fine a set of dimensions, which would lead to a greater and potentially unhelpful number of categories. For example, good factors identifying whether an object should be categorized as a chair could include something with (generally) four legs and something you sit on. Poor factors for making this decision include the material it is made from, or whether it is cushioned. When people use fewer, more abstractable dimensions (especially shared dimensions such as those based on perceptual features), it becomes more likely that their categories will overlap and, so, be more similar to each other.

1.3 Effects of novel labels on categorization

So far, evidence has shown that the use of existing, conventionalized labels influences how people categorize objects. But people frequently encounter novel labels for new or indeed established objects (e.g., names for new technologies), and such novel labels can also influence people's categories. These labels are novel in the sense that they lack a conventionalized entry within the mental lexicon (Jackendoff, 2002), and – as Lupyan, Rakison and McClelland (2007) phrased it – they are 'semantically empty'.

Novel labels actively influence category learning by guiding attention to the relevant perceptual similarities of objects for both infants (Waxman & Markow, 1995) and adults (Lupyan & Casasanto, 2014). Presenting non-word labels (e.g., 'Look at the Timbo!') to infants alongside new objects guided their eye movements to shared features across those objects, so that they increasingly directed their attention more towards shared features than towards dissimilar features, and this led to enhanced category learning (Althaus & Mareschal, 2014; Althaus & Plunkett, 2016).

Lupyan and Casasanto (2014) had adults categorize novel alien-like stimuli into two pre-determined categories using conventionalized labels (i.e., 'smooth headed' vs. 'pointy headed') or non-word labels (i.e., 'fooves' vs. 'crelches'). Participants performed equally well in learning to assign stimuli to the relevant categories (i.e., learning whether an alien should be categorized as a friendly 'foove', or an unfriendly 'crelch') whether they used conventionalized word or non-conventionalized, non-word labels. When labels were sound-symbolically congruent (e.g., smooth-headed aliens to be labelled as 'smooth', or 'foove'), performance for both label types was better than in incongruent trials (e.g., pointy-headed aliens to be labelled as 'smooth', or 'foove'), and also more so than in the control condition, in which participants categorized objects without labels.

Lupyan and Casasanto (2014) argued that, similarly to infants, the presence of novel labels drew attention to the necessary perceptual features (i.e., pointedness vs. smoothness) across objects that supported category learning within individual participants. Their results also showed that – across participants – people came to more effectively associate the same stimuli with the same labels for congruent trials (e.g., people commonly categorized the same aliens as 'fooves'). This is unsurprising, as participants were learning pre-determined categories with pre-assigned labels (i.e., their task was to learn a set of distinctions, including a distinction between two labels and the objects to which they referred) and although these non-words did not have conventionalized meanings, they were constructed to reflect category dimensions through sound symbolism. Nevertheless, their results demonstrate that under some circumstances, labels that lack conventionalized meanings within the mental lexicon (Jackendoff, 2002) can have consistent effects on individuals' categories.

1.4 What can serve as a label?

There appear to be some limits to what can serve as advantageous labels (i.e., labels that influence and facilitate category formation). Lupyan et al. (2007) investigated the efficacy of printed and spoken non-word labels (i.e., *leebish* vs. *grecious*) compared with non-linguistic, location-based cues as labels (i.e., location of the object onscreen) for people learning to categorize novel alien stimuli as 'friendly' or 'unfriendly'. Location-based cues involved the alien moving vertically onscreen to indicate 'where it lived' (e.g., whether the alien came from the 'friendly' part of the planet). Both printed and spoken word labels facilitated category learning, but location-based cues did not.

Lupyan et al. (2007) argued that word labels in their study bolstered category learning by simplifying the distinction between categories. That is, using word labels meant that participants could categorize the objects under a single term (here, *leebish* or *grecious*) that represented the necessary category dimensions and made the distinctions more concrete. In contrast, when they did not use labels, they had to rely on more complex and fuzzy perceptual distinctions (e.g., objects categorized as 'more rounded and smooth' and 'less rounded, with ridges'). They argued that location-based cues did not serve as a label for the existing category dimensions, and so did not facilitate category learning, perhaps because adults prioritize words as referring labels (e.g., nouns as object names), and not facts (i.e., facts such as where an alien lives; Colunga & Smith, 2005).

A further constraint on potential labels is the extent to which they can be sufficiently abstracted to represent the combination of dimensions that make up a category, without being too restricted to certain referents (Lupyan, 2008). Edmiston and Lupyan (2015) argued that linguistic labels (such as *dog* or *guitar*) are advantageous for processing category members because they can be abstracted over all kinds of dog or guitar. In contrast, environmental sounds (such as a dog's bark, or the strum of a guitar) do not produce the same advantages in categorization because they reference specific examples of objects (Lupyan & Thompson-Schill, 2012). For example, if we have three objects from the category of DOG (e.g., a Chihuahua, a Doberman and a Border Collie), the word *dog* can both refer to each object and also group them together. But a particular pitch of bark (e.g., a high-pitched yap) plausibly refers only to the Chihuahua.

Despite location and environmental sound cues not yielding the same benefits for categorization as linguistic, non-word labels, other non-linguistic cues might still be able to affect categorization in a way that increases category coherence across people. Lupyan et al. (2007) predicted that non-linguistic cues could affect categorization, but that this might only be true as long as these cues simplified the distinction between categories and were not applied to objects at a more specific level (i.e., as long as the non-linguistic cue was not treated as another dimension of the object, rather than as a label for the existing dimensions). Non-linguistic cues that meet these criteria should in principle be able to reference category members in the same way as linguistic labels, and accordingly yield the same benefits for categorization.

Therefore – to compare the effects of novel linguistic and non-linguistic labels – the non-linguistic labels, like non-words, should not have conventionalized associations with items, and, therefore, the item categories. For example, colours (despite having existing associations to aspects such as mood, or personality; Ou, Luo, Woodcock & Wright, 2004) should not be readily associated with objects that do not differ by colour. If non-linguistic labels do affect category coherence in the same manner as linguistic non-words, then this effect would provide further support for accounts by which having conventionalized label meanings is not necessary for labelling effects to occur on people's categories.

1.5 Exposure to categories

What is still unknown is whether such novel labels (i.e., linguistic non-words or non-linguistic coloured tags) can influence categorization in a way that increases category coherence in the same way that conventionalized labels do, when people form their own categories (i.e., when they are not learning pre-determined categories, as in Lupyan et al., 2007; Lupyan & Casasanto, 2014). If labelling affects categorization directly by influencing the way in which people attend to object features, then the simple act of using labels during categorization should cause people to develop more similar categories. In other words, pairs and even groups of participants would come to have more similar categories when sorters had used novel labels during categorization than when they had not, even if they were never directly exposed to each other's categories and labels.

However, other work has suggested that exposure to another person's categories (e.g., as part of referential communication) plays a fundamental role in learning how that person conceives and labels particular concepts, and hence in developing categories that are similar to theirs. For example, Markman & Makin (1998) found that communication during paired tasks led to partners having subsequently more similar categories than did pairs who did not share referential communication during the task. As such, they argued that communication is necessary for greater coordination of people's concepts and, thus, categories.

1.6 Current study

Research has shown that when people categorize objects, they are affected by linguistic factors. The availability of conventionalized labels affects the conceptual information that people use to categorize, drawing attention to more consistent and generalizable features in a manner that can lead to greater coherence across people's categories. However, the use of novel linguistic labels has also been shown to affect how people learn pre-determined categories (i.e., categories defined by the experimenters), helping them to develop robust representations that are also based on more consistent and generalizable features. But we do not know whether novel, non-conventionalized labels can affect self-determined categorization across people in a consistent manner, and whether exposure between people is necessary for novel labels to affect category coherence in this way.

If labels do affect categorization directly by influencing which dimensions people select for parsing objects into categories (as was suggested by Waxman & Markow, 1995; Lupyan, 2008), we would expect the categories that people instantiate to be more similar when those categories were instantiated using non-conventionalized labels than when they were not: Using labels would encourage people to select the most abstractable, perceptual dimensions for categorization, and these dimensions should be similar across individuals.

Some theories also that predict if novel labels affect category formation, these labels need not be linguistic since even linguistic novel labels (non-words) do not have conventionalized entries in the mental lexicon. Lupyan et al. (2007) predicted that non-linguistic cues could affect categorization, but that this might only be true as long as these cues simplified the distinction between categories and were not applied to objects at a more specific level (cf. the environmental sounds in Edmiston & Lupyan, 2015). Non-linguistic cues that meet these criteria should in principle be able to reference category members in the same way as linguistic labels, and accordingly yield the same benefits for categorization. As such, we hypothesize that non-linguistic cues such as coloured tags (applied to uncoloured stimuli) could similarly influence people to select more abstractable, perceptual dimensions that work well across a range of objects, and increase people's category coherence.

Lastly, if labelling affects categorization directly by influencing the way in which people attend to object features, then the simple act of using labels during categorization should cause people to develop more similar categories, even if they were never directly exposed to each other's categories and labels. However, other accounts suggest that people might be more likely to develop similar categories to each other when using non-conventionalized labels only if they were able to experience each other's categories (Markman & Makin, 1998; Steels & Belpaeme, 2005). As such, they might only develop more similar categories when novel labels are used in conjunction with exposure to a partner's categories and labels.

We investigated whether two forms of novel label (i.e., linguistic non-words and non-linguistic coloured tags) increased category coherence, and whether the effects were dependent on exposure to a partner's categories. We conducted two experiments in which pairs of participants sorted two sets of geometric stimuli (i.e., morphed, monochromatic triangles that varied in size, shape, angles, and pointedness) into two self-determined categories over ten rounds. They either had exposure to their partner's categories between rounds of sorting, or not.

We did not allow discussion between participants (as this, alongside the use of pre-defined novel labels, might have led to a ceiling effect on category coherence scores), but, instead, allowed exposure (without discussion) to a partner's categories at intervals between sorting. The morphed, randomly-generated triangles were chosen as stimuli because they vary continuously on a small number of perceptual features, and because it is unlikely that people would have strongly existing category boundaries for such items. Participants were given access to the full stimuli space when sorting, as we believed that giving participants access to every item would allow them to more sensibly select category dimensions conducive to the formation of more robust category distinctions based on abstractable, perceptual dimensions.

Firstly, we examined pair coherence, defined as the category coherence for pairs who performed the task together, both with and without exposure to a partner's categories. This measure of pair coherence allowed us to assess the effects of labelling when pairs of sorters were exposed to each other's category systems, or not. If labelling directly influences sorters to pick more abstractable, perceptual dimensions for sorting, then labels might increase pair coherence regardless of exposure (and label type). Alternatively, labelling effects on category coherence might be restricted to circumstances where participants experience each other's categories. If experience of the other's categories plays a role, categories might also become increasingly similar with increasing exposure.

Secondly, we examined group coherence, defined as the category coherence across all participants within an experimental condition. This measure of group coherence allowed us to assess whether there were generalizable effects of labelling and exposure on category coherence across people in general, not just between the pairs who performed the task together. Group coherence therefore reflected whether sorters became more similar to an average way of sorting as a group (i.e., through having labels and having exposure to a partner, or not). Again, if labels directly influenced people's categories, then labels might increase the coherence of people's categories as a group (i.e., despite these participants never being exposed to each other's categories).

Experiment 2

2.1 Methods

2.1.1 Participants. Sixty-four British participants (45 female) were randomly paired to form 32 experimental pairs. Ages ranged from 18 to 27 years (*M* = 19.86, *SD* = 2.56). The University of Edinburgh's Psychology Ethics Committee approved this study.

2.1.2 Stimuli. Stimuli were two sets of 26 monochromatic triangular shapes (henceforth, triangles) which varied in the length of sides (and therefore angles) and in the degree of pointedness of vertices (see Fig. 1; see Appendices H & I for the full sets). Triangles were printed on 3x3 inch cards. The two sets of triangles (Set A and B) each comprised 26 items. The non-word labels were 'TEB' and 'DUP'. These were presented as printed labels alongside each of the stimuli sets for the rounds with labels. Labels remained the same for participants across rounds.

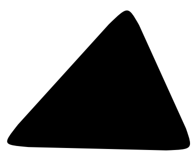


Fig. 1. Example stimulus from Set A.

2.1.3 Design. There were two independent factors: Exposure (between-participants; Exposed vs. Non-exposed) and Labels (within-participants; With-labels vs. No-labels). In the Exposed conditions participants viewed their partner's categories at intervals. In the Non-exposed conditions, participants never saw their partner's categories. For the factor of Labels, participants who had access to non-word labels (With-labels) in block 1 (e.g., rounds 1–5) did not have access to labels (No-labels) in block 2 (e.g., rounds 6–10), and vice versa (see Fig. 2). Set Order (A-B or B-A) was also counterbalanced. A barrier was used to obscure the participants from each other. In the Exposed conditions, this barrier remained in place for the sorting phases, but was removed during intervals so that participants could see each other as well as their categories. In the Non-exposed condition, the barrier was not removed at any point.

Participants were instructed to categorize the stimuli into two categories so that each category contained a minimum of 9 and a maximum of 17 triangles. Our reasoning was that requiring a specific number of triangles per category (e.g., 13 each) could have led to participants 'forcing' triangles into categories that they deemed inappropriate and allowing complete freedom could have led to very small categories.

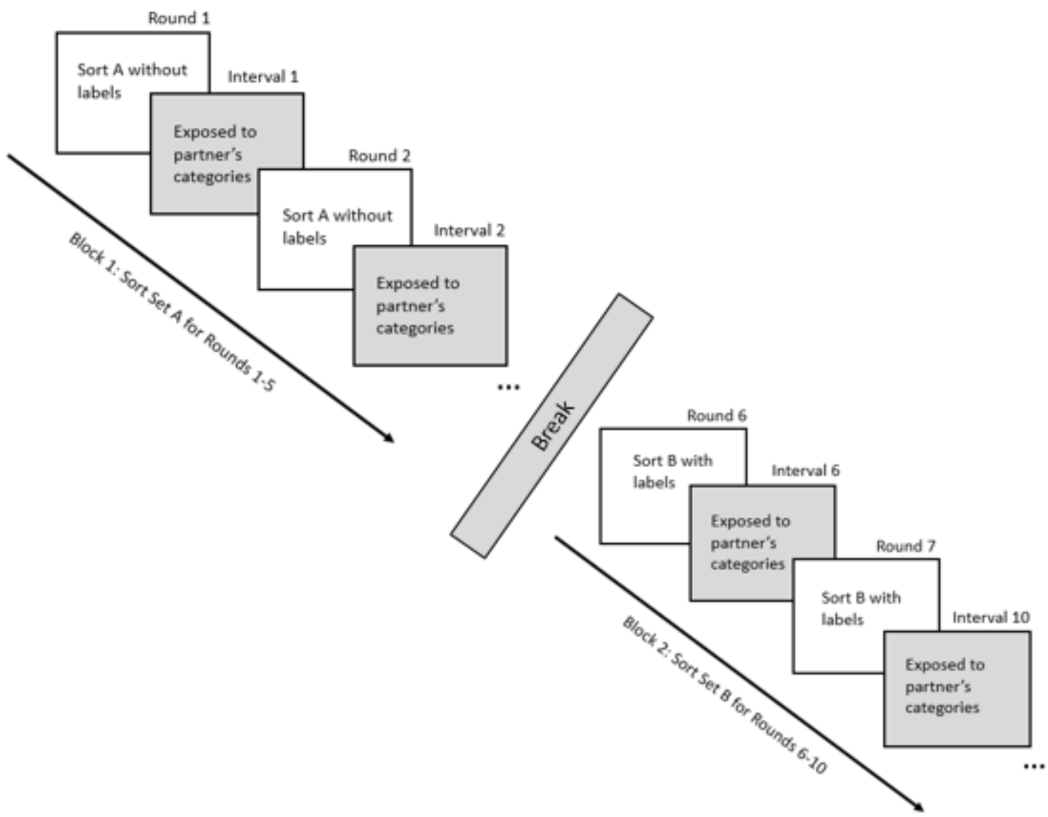


Fig. 2. Example of structure (Exposed); in block 1 participants sorted Set A for 5 rounds without labels, and then in block 2 participants sorted Set B for a subsequent 5 rounds with labels.

2.1.4 Procedure. Participants were seated opposite one another with the barrier in place (see Fig. 3). They were instructed to 'Sort the triangles into two categories in a way that would make sense to you, as well as to another person'. We used these instructions (i.e., rather than giving them an explicit goal of coordinating their categories), as it allowed us to use the same instructions for both the Exposed and Non-exposed conditions (i.e., since explicitly asking participants in the Non-exposed condition to coordinate their categories might be deemed impossible). Participants sorted one set of A or B into two categories, with a time limit of 2 ½ min. All triangles had to be assigned to a category. Participants were free to sort using the same criteria across rounds, or to change the criteria across rounds as they wished. For rounds with labels, printed labels were presented alongside the stimuli cards. Upon receiving the labels participants were told 'You have these two labels to place upon your categories. Place one label on one category each. You choose how to use them. You can move them across the categories between rounds if you wish'.

There was a 30s interval between sets. During this interval, the barrier was removed for participants in the Exposed conditions so that they saw their partner's two categories (and labels if they were on the Labels block), but no communication was allowed; the barrier was not removed for participants in the Non-exposed conditions so that they did not see their partner's categories. Instead, they were asked to use the interval to silently reflect upon how they had categorized the items. Following the interval, the barrier was then replaced for Exposed conditions. After each interval, the sorted cards were collected for recording results and a new deck of stimuli cards was provided; within a block, this would be the same set for 5 rounds. After round 5 (i.e., end of block 1), participants took a short break from the task in which they completed a demographic questionnaire. In block 2, whichever set was not used in block 1 was sorted over rounds 6–10.

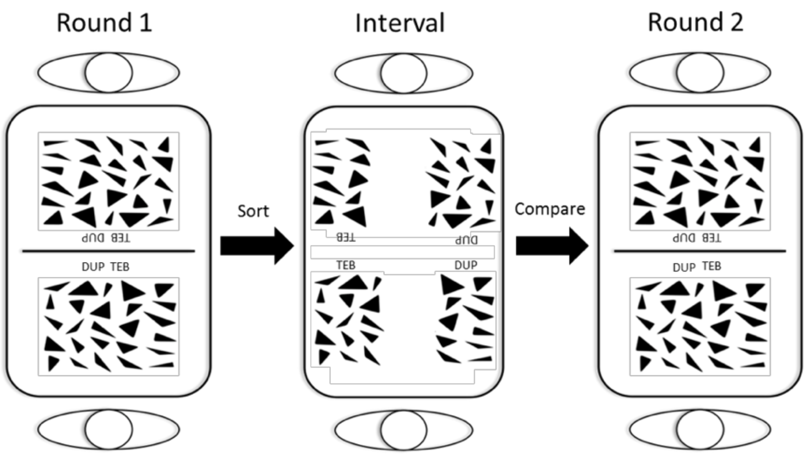


Fig. 3. Example of two rounds of sorting with labels and an interval in the middle (Exposed conditions).

2.2 Results

2.2.1 Pair coherence. Firstly, we investigated the effects of labels and exposure on the 32 pairs of participants who performed the experiment together. For every participant, we coded whether they put each possible pair of items (26x25/2=325 item pairs per round) into the same category (arbitrarily 0 or 1), or not, and used this data to compute a measure of association between participants who had been paired with each other in the experiment (see Fig. 4). If participants placed two items into the same category, that item pair would be coded as 1; if not, the item pair was coded as 0. This resulted in binomial data for 32 participant pairs in total.

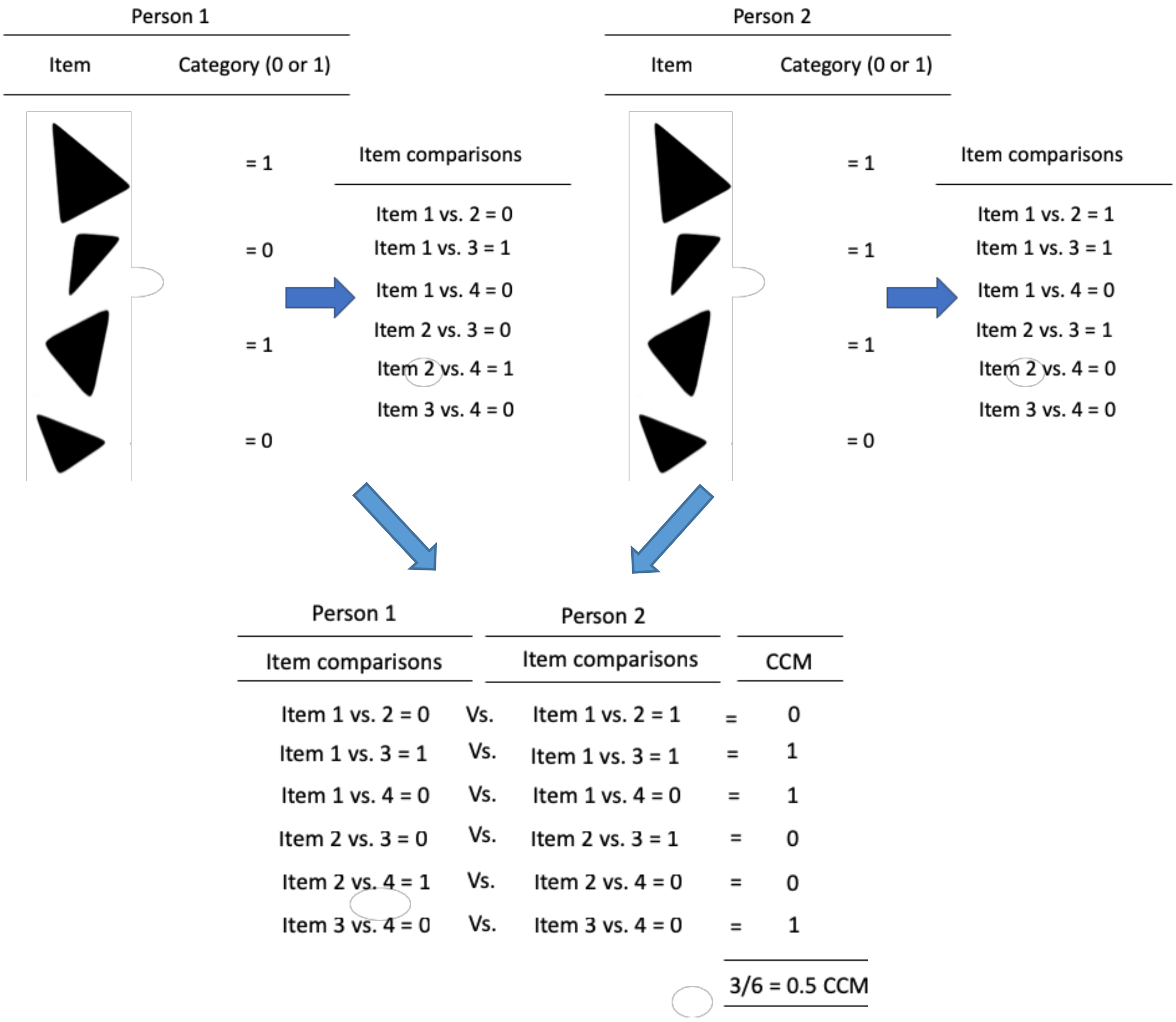


Fig. 4. Example of binomial and proportional scoring methods for two participants using a subset of the items (for the proportional method used in subsequent group analyses, this pair would score 0.5 category coherence).

Descriptive statistics. In order to gain an average of pair coherence across conditions, we summed the total number of item matches between a pair of participants, and divided this by the maximum number of item matches to gain a category coherence score (summarized across Labels and Exposure in Table 1).

Table 1.

Experiment 2, pair coherence: Average category coherence score (SD) by Labels and Exposure.

Condition	Coherence	SD
Non-exposed; No labels	0.296	0.061
Exposed; No labels	0.329	0.080
Non-exposed; With labels	0.310	0.071
Exposed; With labels	0.332	0.067

Generalized logit mixed-effects models (GLMM) analysis. We analysed the binomial pair coherence results using a GLMM approach in in R 3.2.1 (R Core Team, 2015), with the lme4 package, version 1.1-8 (Bates, Maechler, Bolker & Walker, 2014). This approach allowed us to account for random variance due to differences between participant pairs and by round. The threshold for statistical significance was set at $|p| < .05$ (Baayen, 2008). A backwards, stepwise elimination approach was used to select factors for the final model, using likelihood ratio tests to compare models.

The initial model included Labels and Exposure as fixed effects, as well the interaction between these factors. The initial model included random slopes and intercepts for Labels and Round by Participant Pair (i.e., the pairs of participants performing the task), and for Exposure and Label order by Round. This initial model was a significantly better fit of the data than the null model ($X^2(9) = 396.35$, $p < .001$). Removing the interaction term for Labels and Exposure did not significantly reduce model fit over the initial model ($p > .05$). Removing Labels from the model did not significantly reduce model fit ($p > .05$). Thus, the model of best fit included Exposure as a fixed effect (see Tables 2 & 3). Under this model, there was a significant effect of Exposure, such that pairs of participants who were exposed to each other's categories had greater pair coherence than those who were not exposed to each other's categories ($\beta = 0.08$, SE = 0.04, $p < .05$).

Table 2.

Pair coherence analysis of Experiment 2: Beta, standard errors, z and p-values for fixed effects on category coherence. Model fit by REML.

Fixed Effects	**β**	**S.E.**	**z**	**p**
Intercept	-0.77	0.04	-18.03	<.001
Exposure	0.08	0.04	2.05	<.04

Table 3.

Pair coherence analysis of Experiment 2: Variance for random effects. Model fit by REML.

Random Effects		
Pair	Intercept	0.04
	Labels	0.02
	Round	0.02
	Labels*Round	0.02
Round	Intercept	0.01e-01
	Exposure	0.01e-02
	Label order	0.01e-01
	Exposure*Label order	0.01e-02

No. of Observations = 104000.

2.2.2 Group coherence. We investigated the effect of labels and exposure at the group-level by comparing group coherence between every possible pair of participants within a condition within Experiment 2. That is, we took every individual in Experiment 2 from the Exposed condition and compared their categories to every other person from that condition, and then took every individual in Experiment 2 from the Non-exposed condition and compared their categories to every other person from that condition. It is crucial to note that Exposure for group coherence refers to the fact that participants were or were not exposed to one other person's categories, or not, but that for this analysis their categories were compared to the categories of people that they were not partnered with in the experiment. That is, we investigated the effects of being exposed to a single person's categories on group coherence across all sorters within a condition. From this, we were able to compare group coherence on the basis of Labels, Exposure, Label order and Round in all possible pairings of participants in a condition. If our predictions held regarding the generalizable effects that novel labels and exposure to a person's categories can have on group coherence, then there would be an effect of Labels and Exposure even for the non-interacting pairs which formed the basis of this group analysis.

Calculating coherence across all pairs. In order to be make multiple comparisons for every possible pair, we implemented the cultural consensus model (CCM) (Romney, Weller & Batchelder, 1986; Malt et al., 1999; Ameel et al., 2005; White, Malt & Storms, 2016). For every participant, we again coded whether they put each possible pair of items (26x25/2=325 item pairs per round) into the same category, or not, and used this data to compute a measure of association between participants. Again, if participants placed two items into the same category, that item pair would be coded as 1; if not, the item pair was coded as 0. We used this to calculate a proportional score between 1 and 0 and used these scores into our group coherence analysis (i.e., in contrast to the binomial data used in the pair coherence analysis). We calculated proportional scores for every possible pair of participants within Experiment 2 for Exposure, Label order, Labels and Round, and then averaged the resulting proportions by participant (see Fig. 4 for proportional scoring). This resulted in an average association score by round for each of the 64 participants.

Descriptive Statistics. Average group coherence scores by all possible pairs for each condition, within Experiment 2, are summarized across Labels and Exposure in Tables 4-7. In addition, individual data points (average by participant) are plotted across the conditions as bee-swarm plots (see Figures 5-8).

Table 4.

Experiment 2, group coherence: Average category coherence score (SD) for Non-exposed and No-labels.

Exposure	Labels	Round	Average CCM	SD
Non-exposed	No-labels	1	0.269	0.043
Non-exposed	No-labels	2	0.283	0.058
Non-exposed	No-labels	3	0.279	0.040
Non-exposed	No-labels	4	0.262	0.016
Non-exposed	No-labels	5	0.278	0.043
Non-exposed	No-labels	6	0.255	0.016
Non-exposed	No-labels	7	0.263	0.010
Non-exposed	No-labels	8	0.297	0.044
Non-exposed	No-labels	9	0.266	0.014
Non-exposed	No-labels	10	0.265	0.015

Table 5.

Experiment 2, group coherence: Average category coherence score (SD) for Non-exposed and With-labels.

Exposure	Labels	Round	Average CCM	SD
Non-exposed	With-labels	1	0.270	0.027
Non-exposed	With-labels	2	0.266	0.023
Non-exposed	With-labels	3	0.258	0.011
Non-exposed	With-labels	4	0.290	0.049
Non-exposed	With-labels	5	0.254	0.013
Non-exposed	With-labels	6	0.274	0.018
Non-exposed	With-labels	7	0.272	0.031
Non-exposed	With-labels	8	0.260	0.015
Non-exposed	With-labels	9	0.278	0.050
Non-exposed	With-labels	10	0.330	0.091

Table 6.

Experiment 2, group coherence: Average category coherence score (SD) for Exposed and No-labels.

Exposure	Labels	Round	Average CCM	SD
Exposed	No-labels	1	0.276	0.030
Exposed	No-labels	2	0.283	0.020
Exposed	No-labels	3	0.298	0.031
Exposed	No-labels	4	0.293	0.038
Exposed	No-labels	5	0.276	0.022
Exposed	No-labels	6	0.310	0.061
Exposed	No-labels	7	0.290	0.039
Exposed	No-labels	8	0.284	0.031
Exposed	No-labels	9	0.312	0.069
Exposed	No-labels	10	0.342	0.080

Table 7.

Experiment 2, group coherence: Average category coherence score (SD) for Exposed and With-labels.

Exposure	Labels	Round	Average CCM	SD
Exposed	With-labels	1	0.295	0.053
Exposed	With-labels	2	0.281	0.040
Exposed	With-labels	3	0.300	0.046
Exposed	With-labels	4	0.288	0.054
Exposed	With-labels	5	0.310	0.062
Exposed	With-labels	6	0.283	0.058
Exposed	With-labels	7	0.302	0.044
Exposed	With-labels	8	0.284	0.045
Exposed	With-labels	9	0.280	0.028
Exposed	With-labels	10	0.311	0.051

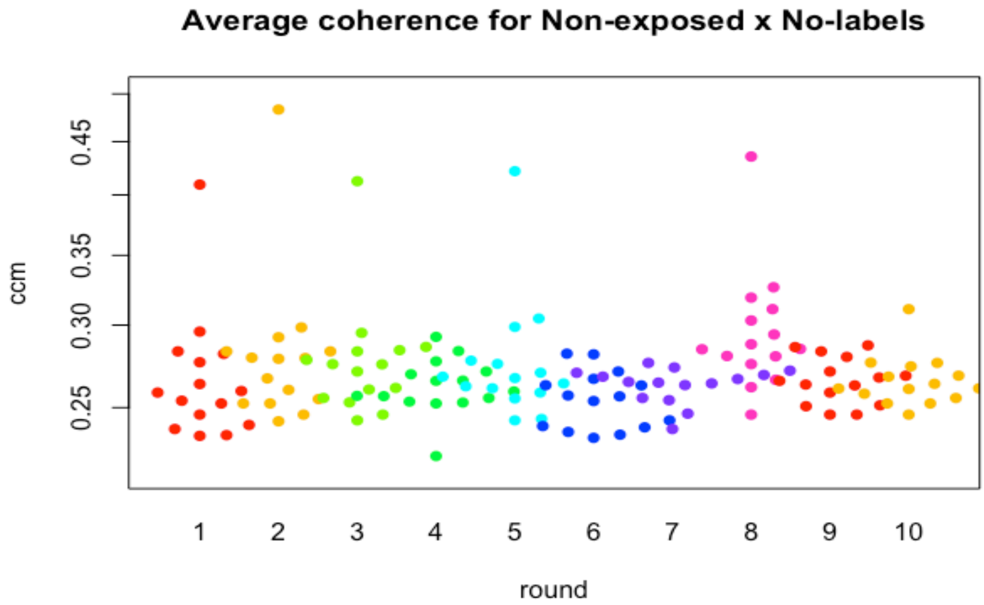


Fig. 5. Experiment 2; Average category coherence (CCM) by Round (1-10) for participants in the Non-exposed x No-labels condition.

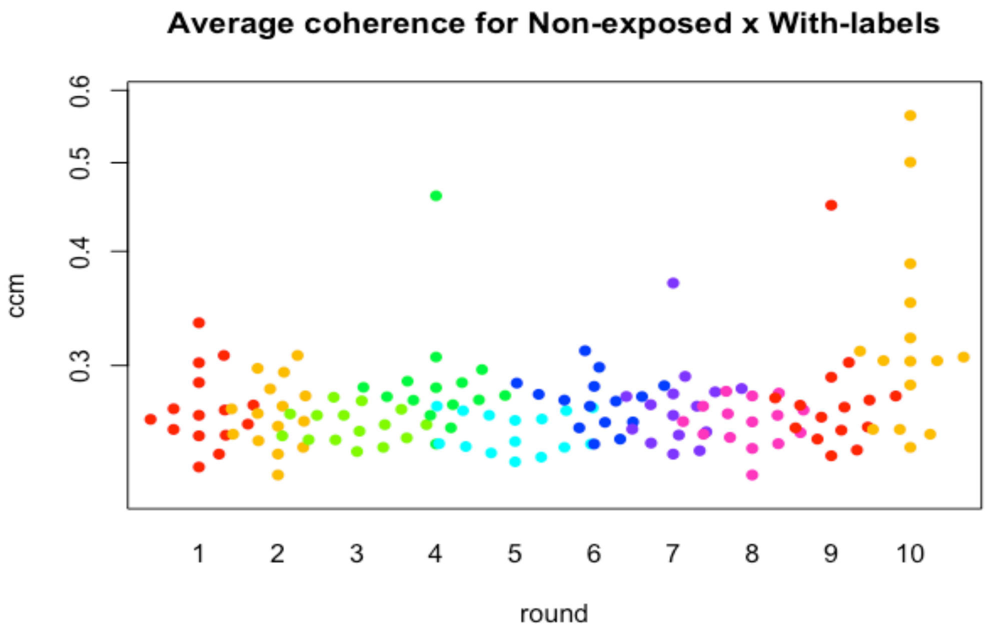


Fig. 6. Experiment 2; Average category coherence (CCM) by Round (1-10) for participants in the Non-exposed x With-labels condition.

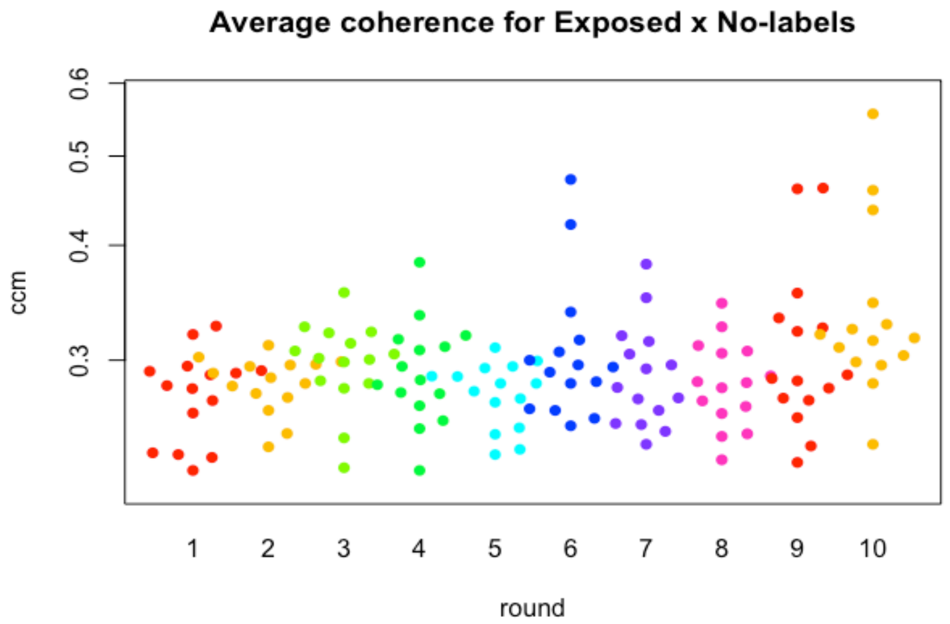


Fig. 7. Experiment 2; Average category coherence (CCM) by Round (1-10) for participants in the Exposed x No-labels condition.

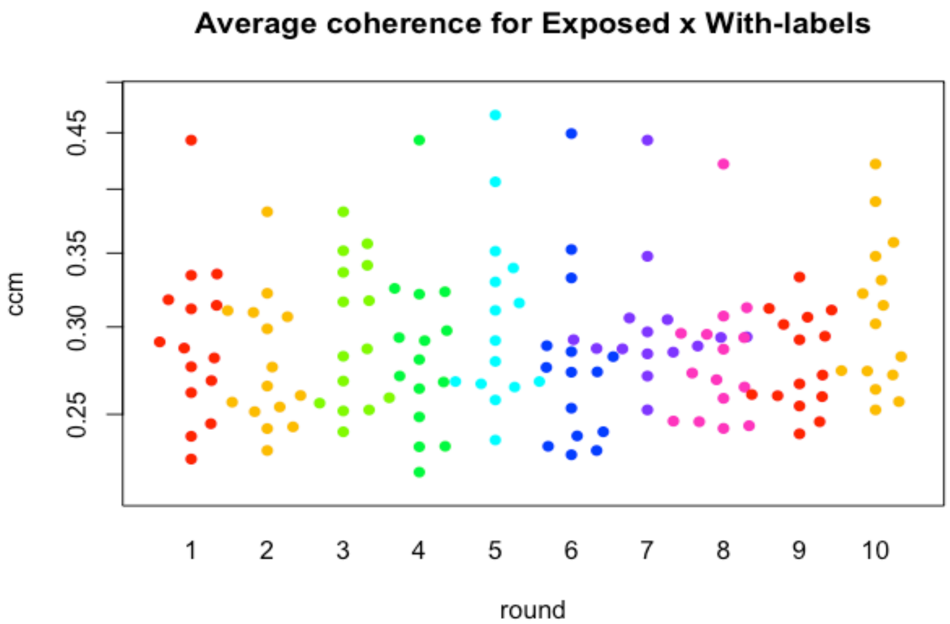


Fig. 8. Experiment 2; Average category coherence (CCM) by Round (1-10) for participants in the Exposed x With-labels condition.

Linear mixed-effects models (LME) analysis. To test for the effects of Labels and Exposure on group coherence, data were analyzed in R with a linear mixed-modeling approach again using the lme4 package (Bates et al., 2014). This approach allowed us to account for random variance due to differences between participants and by round. The threshold for statistical significance was set at $|t| > 2$ (Baayen, 2008). A backwards, stepwise elimination approach was used to select factors for the final model, using likelihood ratio tests to compare models.

In order to normalize the sampling distribution of the CCM scores, they were Z-transformed with $Z=0.5*\ln[(1+r)/(1-r)]$. These Z-transformations were then analysed using the lme4 package in R. The initial model included Labels and Exposure as fixed effects, as well the interaction between these factors. The initial model included random slopes and intercepts for Labels and Round by Participant, and for Exposure and Label order by Round. This full model was a significantly better fit than the null model ($X^2(9) = 53.51$, $p < .001$). Removing the interaction term for Labels and Exposure did not significantly reduce model fit over the initial model ($p > .05$). Removing Labels from the model did not significantly reduce model fit ($p > .05$). Thus, the model of best fit included Exposure as a fixed effect (see Tables 8 & 9). Under this model, there was a significant effect of Exposure, such that pairs in the Exposed condition had greater group coherence than pairs in the Non-exposed condition ($\beta = 0.25$, SE = 0.07, $t = 3.36$).

Table 8.

Group coherence analysis of Experiment 2: Beta, standard errors and t-values for fixed effects on category coherence Z-score. Model fit by REML.

Fixed effects	***β***	**S.E.**	***t***
Intercept	-0.02	0.10	-0.18
Exposure	0.25	0.07	3.36

Table 9.

Group coherence analysis of Experiment 2: Variance and residual for random effects. Model fit by REML.

Random Effects		
Participant	Intercept	0.20
	Labels	0.02
	Round	0.02
	Labels*Round	0.13
Round	Intercept	0.04
	Exposure	0.03e-01
	Label order	0.01
	Exposure*Label order	0.04
Residual		0.48

No. of Observations = 640.

2.3 Summary of Experiment 2

Despite the results of Experiment 1 which suggested a general effect of labelling on categorization, we found no effect of having labels on pair or group coherence, but we found a significant effect exposure on both pair and group coherence, such that having exposure to a partner led to greater coherence, than did sorting without exposure. Overall, these results suggest that exposure to a partner's categories yields benefits for both pair and category coherence.

Experiment 3

In Experiment 3, we further investigated the effects of Labels and Exposure on category coherence; specifically, we examined whether non-linguistic labels can affect categorization and category coherence, despite the lack of effects for linguistic non-word labels in Experiment 2. We therefore used the same design as Experiment 2, but replaced the non-word labels with non-linguistic, coloured tags.

3.1 Methods

3.1.1 Participants. Sixty-four new British participants (48 female) formed 32 experimental pairs**.** Ages ranged from 18 to 30 years (*M* = 20.39 years, *SD* = 2.87).

3.1.2 Stimuli. Stimuli were the same sets (A & B) of triangular shapes used in Experiment 2 and they were presented in an identical manner**.** The coloured tags of blue and yellow (i.e., 5PB and 5Y respectively on the Munsell colour system; Munsell, 1912) were presented alongside the stimuli sets in the With-labels condition and were absent in the No-labels condition. These colours were chosen because they are frequent, focal colours that we judged not to have any strong semantic or emotional associations to the stimuli. Labels remained the same for participants across rounds.

3.1.3 Design. The design was identical to Experiment 2, such that again there were two independent factors: Exposure (between-pairs; Exposed vs. Non-exposed) and Labels (within-pairs; With-labels vs. No-labels). We also included the factor of Label order (between-pairs: Labels 1st vs. Labels 2nd), to account for potential differences in category coherence caused by the counterbalancing of label presentation.

3.1.4 Procedure. The procedure was identical to Experiment 2, except for the type of labels provided on With-labels rounds.

3.2 Results

3.2.1 Pair coherence. Again, we investigated the effects of labels and exposure on a further 32 pairs of participants who performed the experiment together, leading to 325 observations per participant per round. If participants in a pair placed two items into the same category, that item pair would be coded as 1; if not, the item pair was coded as 0. This resulted in binomial data for 32 participant pairs in total.

Descriptive statistics. To gain an average of pair coherence across conditions, we again summed the total number of item matches between a pair of participants, and divided this by the maximum number of item matches (summarized across Labels and Exposure in Table 10).

Table 10.

Experiment 3, pair coherence: Average category coherence score (SD) by Labels and Exposure.

Condition	Coherence	SD
Non-exposed; No labels	0.302	0.068
Exposed; No labels	0.323	0.072
Non-exposed; With labels	0.325	0.075
Exposed; With labels	0.329	0.072

GLMM analysis. As in Experiment 2, we analysed the binomial data for the 32 pairs using a GLMM approach, and accounting for random variance due to differences between participant pairs and by round. The initial model included Labels and Exposure as fixed effects, as well the interaction between these factors. The initial model included random slopes and intercepts for Labels and Round by Participant Pair, and for Exposure and Label order by Round. This initial model was a significantly better fit of the data than the null model ($X^2(9) = 375.74$, $p < .001$). Removing the interaction term for Labels and Exposure did not significantly reduce model fit over the initial model ($p > .05$). Removing

Labels from the model did not significantly reduce model fit ($p > .05$). Thus, the model of best fit included Exposure as a fixed effect (see Tables 11 & 12). However, under this model there was not a significant effect of Exposure ($p > .05$).

Table 11.

Pair coherence analysis of Experiment 3: Beta, standard errors, z and p-values for fixed effects on category coherence. Model fit by REML.

Fixed effects	**β**	**S.E.**	**z**	**p**
Intercept	-0.77	0.05	-17.00	<.001
Exposure	0.05	0.04	1.12	0.26

Table 12.

Pair coherence analysis of Experiment 3: Variance for random effects. Model fit by REML.

Random Effects		
Pair	Intercept	0.05
	Labels	0.01
	Round	0.01
	Labels*Round	0.02
Round	Intercept	0.08e-02
	Exposure	0.07e-04
	Label order	0.02e-01
	Exposure*Label order	0.07e-03

No. of Observations = 104000.

3.2.2 Group coherence. Again, we investigated the effect of labels at the group-level by comparing group coherence between every possible pair of participants within a condition within Experiment 3, in the same manner as in Experiment 2.

Calculating coherence across all pairs. As in Experiment 2, CCM was used to calculate coherence between every possible pair of participants within a condition. We computed proportional group coherence scores for every possible pair of participants within Experiment 3 for Exposure, Label order, Labels and Round, and averaged this by participant. This resulted in an average association score by round for each of the 64 participants.

Descriptive Statistics. Average group coherence scores by all possible pairs for each condition, within Experiment 3, are summarized across Labels and Exposure in Tables 13-16. In addition, individual data points (average by participant) are plotted across the conditions as bee-swarm plots (see Figures 9-12).

Table 13.

Experiment 3, group coherence: Average category coherence score (SD) for Non-exposed and No-labels.

Exposure	Labels	Round	Average CCM	SD
Non-exposed	No-labels	1	0.275	0.024
Non-exposed	No-labels	2	0.291	0.035
Non-exposed	No-labels	3	0.283	0.026
Non-exposed	No-labels	4	0.287	0.034
Non-exposed	No-labels	5	0.301	0.033
Non-exposed	No-labels	6	0.265	0.017
Non-exposed	No-labels	7	0.267	0.020
Non-exposed	No-labels	8	0.264	0.015
Non-exposed	No-labels	9	0.269	0.017
Non-exposed	No-labels	10	0.281	0.030

Table 14.

Experiment 3, group coherence: Average category coherence score (SD) for Non-exposed and With-labels.

Exposure	Labels	Round	Average CCM	SD
Non-exposed	With-labels	1	0.275	0.021
Non-exposed	With-labels	2	0.264	0.024
Non-exposed	With-labels	3	0.273	0.024
Non-exposed	With-labels	4	0.276	0.037
Non-exposed	With-labels	5	0.264	0.018
Non-exposed	With-labels	6	0.274	0.019
Non-exposed	With-labels	7	0.272	0.022
Non-exposed	With-labels	8	0.262	0.023
Non-exposed	With-labels	9	0.264	0.021
Non-exposed	With-labels	10	0.272	0.022

Table 15.

Experiment 3, group coherence: Average category coherence score (SD) for Exposed and No-labels.

Exposure	Labels	Round	Average CCM	SD
Exposed	No-labels	1	0.283	0.019
Exposed	No-labels	2	0.312	0.050
Exposed	No-labels	3	0.275	0.023
Exposed	No-labels	4	0.318	0.045
Exposed	No-labels	5	0.283	0.027
Exposed	No-labels	6	0.272	0.021
Exposed	No-labels	7	0.276	0.026
Exposed	No-labels	8	0.270	0.019
Exposed	No-labels	9	0.274	0.023
Exposed	No-labels	10	0.276	0.023

Table 16.

Experiment 3, group coherence: Average category coherence score (SD) for Exposed and With-labels.

Exposure	Labels	Round	Average CCM	SD
Exposed	With-labels	1	0.290	0.038
Exposed	With-labels	2	0.275	0.040
Exposed	With-labels	3	0.279	0.035
Exposed	With-labels	4	0.287	0.040
Exposed	With-labels	5	0.267	0.020
Exposed	With-labels	6	0.294	0.035
Exposed	With-labels	7	0.294	0.033
Exposed	With-labels	8	0.284	0.031
Exposed	With-labels	9	0.298	0.038
Exposed	With-labels	10	0.284	0.032

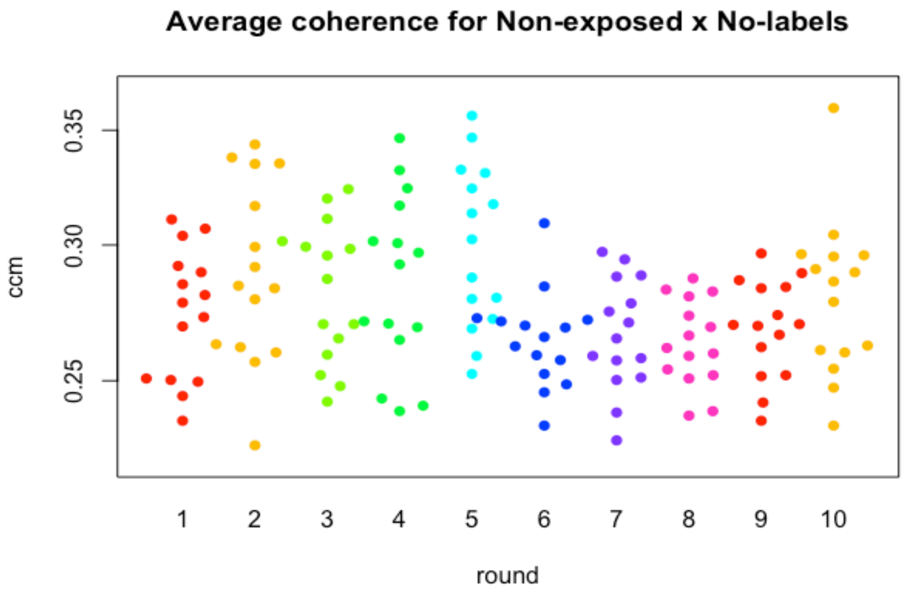


Fig. 9. Experiment 3; Average category coherence (CCM) by Round (1-10) for participants in the Non-exposed x No-labels condition.

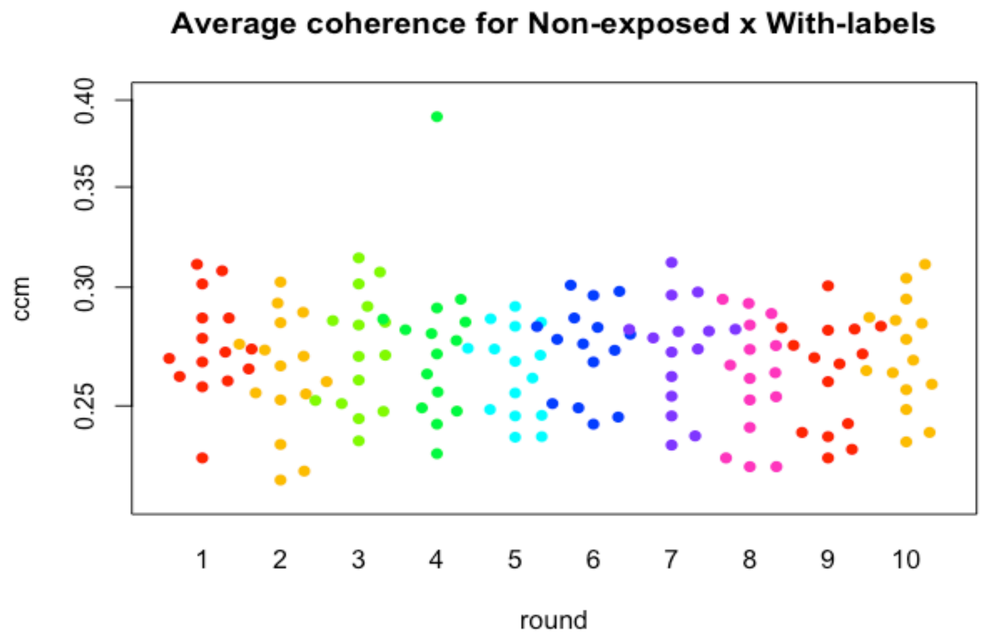


Fig. 10. Experiment 3; Average category coherence (CCM) by Round (1-10) for participants in the Non-exposed x With-labels condition.

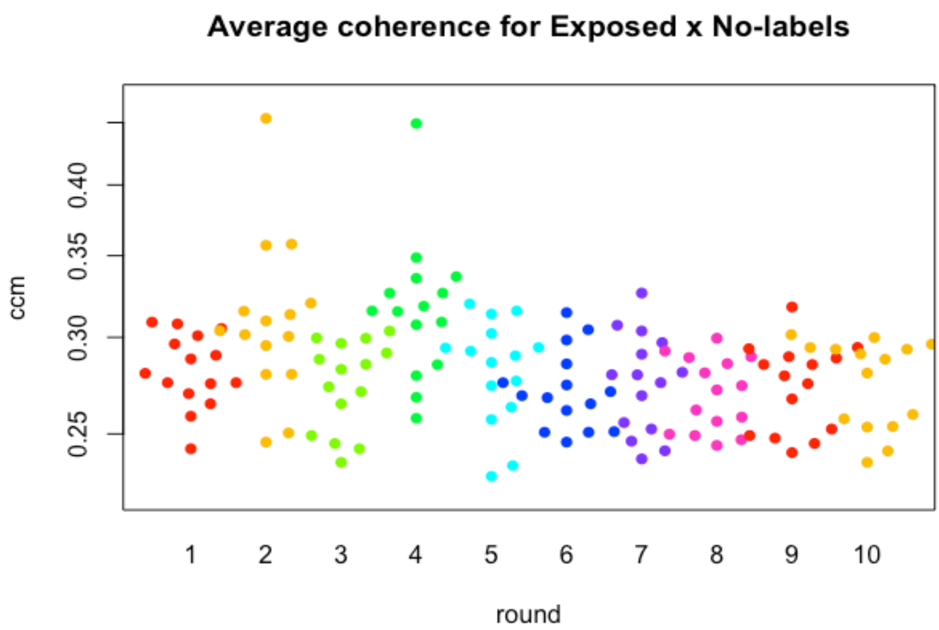


Fig. 11. Experiment 3; Average category coherence (CCM) by Round (1-10) for participants in the Exposed x No-labels condition.

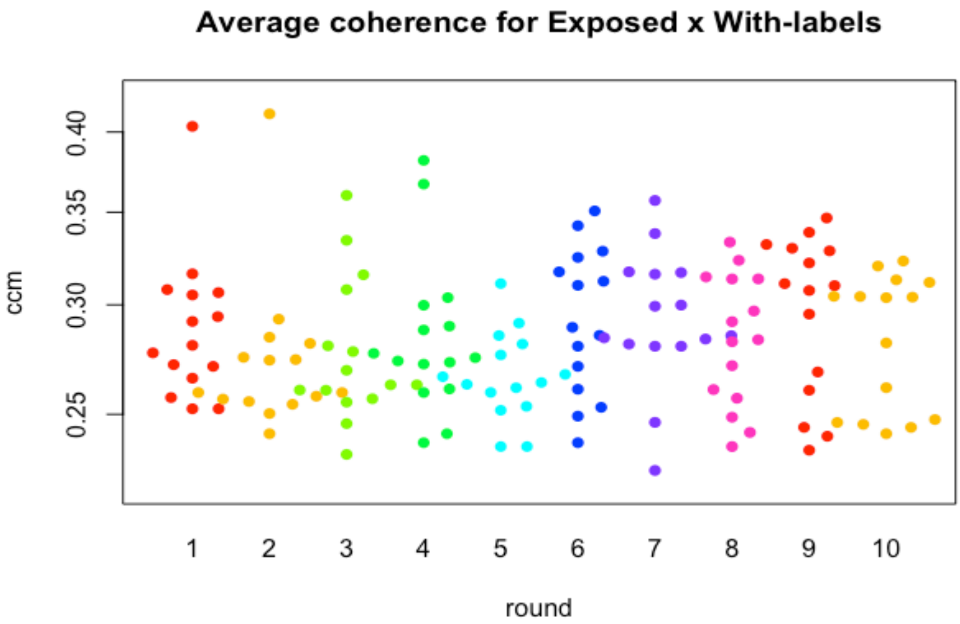


Fig. 12. Experiment 3; Average category coherence (CCM) by Round (1-10) for participants in the Exposed x With-labels condition.

LME analysis. As in the group-level analysis for Experiment 2, proportional group coherence scores were Z-transformed and analysed using an LME approach. The initial model included Labels and Exposure as fixed effects, as well the interaction between these factors. The initial model included random slopes and intercepts for Labels and Round by Participant, and for Exposure and Label order by Round. This full model was a significantly better fit than the null model ($X^2(9) = 106.41$, $p < .001$). Removing the interaction term for Labels and Exposure did not significantly reduce model fit over the initial model ($p > .05$). Removing Labels from the model did not significantly reduce model fit ($p > .05$). Thus, the model of best fit included Exposure as a fixed effect (see Tables 17 & 18). Under this model, there was a significant effect of Exposure, such that pairs in the Exposed condition had greater group coherence than pairs in the Non-exposed condition (β= 0.17, SE = 0.08, t = 2.18).

Table 17.

Group coherence analysis of Experiment 3: Beta, standard errors and t-values for fixed effects on category coherence Z-score. Model fit by REML.

Fixed effects	***β***	**S.E.**	***t***
Intercept	-0.49	0.08	< .001
Exposure	0.17	0.08	2.18

Table 18.

Group coherence analysis of Experiment 3: Variance and residual for random effects. Model fit by REML.

Random Effects		
Participant	Intercept	0.25
	Labels	0.17
	Round	< .001
	Labels*Round	< .001
Round	Intercept	0.02
	Exposure	0.02
	Label order	0.04
	Exposure*Label order	0.03e-01
Residual		0.48

No. of Observations = 640.

3.3 Summary of Experiment 3

Similarly to the non-linguistic labels used in Experiment 2, we found no effect of having labels on pair or group coherence. In contrast to Experiment 2, we also found no effect of having exposure to a partner's categories on pair coherence. However, we found a significant effect of exposure on group coherence, such that having exposure to a partner led to greater coherence, than did sorting without exposure. Overall, these results suggest that exposure to a partner's categories yields benefits for both pair and category coherence. Overall, these results suggest that both linguistic and non-linguistic labels are not always beneficial to increased category coherence across sorters. Instead, the results emphasize the importance of exposure to another person's way of categorizing in producing categories that are more similar to other people's categories.

General Discussion

In two experiments, pairs of participants sorted abstract shape stimuli into two self-determined categories with or without novel labels (i.e., non-words or coloured tags), and with and without exposure to a partner's categories. Despite the results of Experiment 1 suggesting that there may be a general, arbitrary effect of labelling on categorization, there was no significant effect of labels in any of the analyses across Experiments 2 or 3. Instead, there was a significant effect of having exposure to a partner's categories for both pair and group coherence in Experiment 2, and for group coherence in Experiment 3. These results suggest that: a) novel labels do not always benefit category coherence between sorters, and b) that exposure to another sorter's way of categorization may increase coherence between not only the pairs that were exposed to each other's categories, but also across sorters as a group.

4.1 No Effect of Labels

4.1.1 Pair coherence. Previous research has shown that using conventionalized linguistic labels affects how people conceptualize and categorize objects and can increase the similarity of their categories (Markman & Makin, 1998). Other research has shown that novel, linguistic labels affect how infants and adults learn pre-determined categories (Althaus & Mareschal, 2014; Althaus & Plunkett, 2016; Lupyan et al., 2007; Lupyan & Casasanto, 2014), and that these effects can sometimes be consistent across people. One mechanism by which this might occur is that label use supports the selection of more abstractable category dimensions (Waxman & Markow, 1995; Lupyan, 2008). The way in which this might occur is that labels, among other things (like exposure to others' categories), could influence people to select perceptual bases for dimensions, instead of more individualistic dimensions, and these perceptual bases tend not to vary as much across humans. In other words, labels could lead to greater category coherence because they help people to identify appropriate category dimensions, which are also likely to be appropriate for other people.

In this line of thought, our findings from Experiment 1 suggested that novel labels could increase the coherence of people's categories in a way that was arbitrary in that it was not linked to the content of the label itself (i.e., pairs became more similar in how they sorted items regardless of whether they used the labels in specifically the same way as each other). As such, it appeared that there might be general effects of labelling in which the act of labelling itself directly affected the dimensions people selected for their categories - influencing them to pick the same dimensions as each other - and, so, increasing the similarity of their categories. We, therefore, ran Experiments 2 and 3 to investigate whether there were direct effects of labelling on categorization and category coherence, with and without exposure to other sorters' categories.

Despite our predictions following the results of Experiment 1, both Experiments 2 and 3 did not show an effect of labels on category coherence and there was also no interaction of labels with exposure. There are several differences between Experiment 1 and Experiments 2 and 3 that might explain why labels were not beneficial to coherence in the latter two experiments, but crucially there is the difference in how we manipulated exposure to a partner across these experiments. In Experiment 1, all participants apart from those in the control condition, were exposed to each other at the intervals between sorting. This means that participants in the novel, non-word labels condition of Experiment 1 are comparable to participants in the Exposed conditions, but not those in the Non-exposed conditions, of Experiments 2 and 3. As such, it is possible that participants in Experiment 1 only showed greater coherence in the labels condition (than in the dialogue condition), since they had exposure to their partner's categories at the intervals. As for why participants in Experiments 2 and 3 did not show greater coherence with labels and with exposure, it is possible that the exposure to categories itself was more crucial than the use of labels in this particular experimental set-up.

4.1.2 Group coherence. Similarly to the results for pair coherence, both linguistic non-words and non-linguistic colour tags showed no effect on category coherence, regardless of whether participants were exposed to their partner or not. This is expected given the lack of effects for pair coherence, since if labels do not affect the similarity of sorters who are exposed to one another (i.e., within an experimental pair) – then, labels are also unlikely to increase the similarity of sorters as a group. Therefore, our results within this paradigm do not support accounts such as Lupyan's (2008), in which labels directly affect which dimensions we select for sorting, such that we select more abstractable and robust dimensions that other sorters are likely to also select when categorizing with labels. Instead, our results suggest an importance of exposure to another sorter's categories in increasing category coherence.

4.2 Effects of Exposure

4.2.1 Pair coherence. For pairs who performed the task together, having exposure to a partner's categories increased the similarity of their categories for non-word labels (Experiment 2), but not for coloured tags (Experiment 3). It must be noted that exposure in our task meant exactly that – being able to see a partner's categories, but not being able to actively discuss them and, therefore, not being able to use referential communication (as was the case in prior work; e.g., Markman & Makin, 1998). As such, our results do not provide evidence that communication leads to coordination between pairs per se, but that simply being able to see a partner's categories can lead to greater category coherence between pairs, even without interaction.

Results for Experiment 2, therefore, support the importance of exposure to another person's categories in developing categories that are similar to theirs. Accounts that argue that interaction is necessary for coordination also suggest that people might develop more similar categories to each other when using non-conventionalized labels, if they are able to experience each other's categories than if they were not able to, since using shared labels in dialogue has been shown to improve coordination (e.g., Clark & Brennan, 1991; Markman & Makin, 1998). However, we found no interaction between the effects of labels and having exposure in Experiments 2 or 3. This suggests that the mechanisms by which labels and exposure can affect categories do not always interact in a way that increases category coherence.

In contrast to Experiment 2, results for Experiment 3 did not show an effect of having exposure to a partner's categories on pair coherence. As such, it seems that non-linguistic coloured tags did not yield a benefit for category coherence between pairs of sorters who performed the task together.

4.2.2 Group coherence. Both linguistic non-words and non-linguistic coloured tags demonstrated a significant effect on category coherence, leading to greater coherence across sorters as a group. It is again important to reiterate the difference between exposure for pair coherence and exposure for group coherence. For pair coherence, the effect of exposure reflects the direct effect of seeing a partner's categories on the similarity of categories between an individual and that specific partner. In contrast − for group coherence − exposure reflects the effect of seeing one partner's categories, on the coherence of an individual's categories with every other person within the same condition (i.e., despite the fact that the individual was never exposed to these people's categories during the task). As such, exposure for group coherence relates to a more generalizable effect of exposure to one partner's categories on a person's category coherence with other people that they have not been exposed to. This group coherence therefore reflected whether sorters became more similar to an average way of sorting as a group.

The effects of exposure at the group level might, therefore, also relate to the selection of more generalizable and abstractable dimensions – that is, dimensions for categorization that are deemed acceptable across two individuals, and which therefore might also be dimensions most likely to be deemed acceptable across a group of people. This is especially so when those dimensions are based on features shared across people, such as common, perceptual features (Rosch & Mervis, 1975; Johnson, 1987). And by being exposed to a partner, we are more likely to stick to such dimensions across the task and, thus, develop categories that are more similar to other people's (i.e., people who were exposed to another person's categories, but to whose categories we have not been exposed). Relating this to the development of conventions, it is possible that – since linguistic conventions (such as reference usage, Clark & Brennan, 1991) are acquired and maintained mainly through one-on-one interactions (i.e., conversations) – the same mechanisms governing conventions in one-on-one interactions might pertain to some extent to the broader, social context (i.e., how groups develop conventions, Garrod & Doherty, 1994).

4.3 Conclusion

Novel labels did not increase category coherence between pairs of participants regardless of whether they saw each other's groupings, for both linguistic and non-linguistic novel labels. This may have been a result of the specific paradigm we used, in which participants had limited time to categorize a range of items over several rounds, while attempting to understand how a specific partner might categorize the same items, both with and without exposure to the said partner. Instead, our results overall support an account in which exposure to others' categories is sometimes necessary for category coherence, because it suggests that people generally needed to see each other's categories in order to sort items more similarly. The only instance in which this was not the case was for non-linguistic coloured tags in pair coherence in Experiment 3, in which there was no effect of having exposure on the similarity of partners' categories. And, despite this, having exposure to a partner did increase group coherence in Experiment 3. As such, we argued that having exposure to other people's ways of sorting might relate to the selection of more generalizable and abstractable dimensions – that is, dimensions for categorization that are deemed acceptable across two individuals, and which therefore might be those dimensions most likely to be deemed acceptable across a group of people.

Chapter 4

Novel labels increase category coherence, but only within contexts that require coordination[1]

Labels reflect the way that we categorize, but also allow us to communicate and share categories. People sometimes develop more similar categories when they use novel labels to categorize, than when they do not use labels, in categorization tasks. However, we do not know whether the category coherence yielded by novel labels in these tasks is dependent on the need for coordination between partners. To address this, we had participants individually categorize images of mountains with or without novel labels, and with or without a coordinative context. Experiment 4 (*non-coordinative*) did not demonstrate an effect of labels on category coherence across sorters. Experiment 5 (*coordinative*) showed greater category coherence for participants who sorted using novel labels, than those who sorted without labels. These results provide evidence that novel labels can affect categorization in a way that increases the potential for people's categories to overlap. However, the null results of the combined data (contrasting Experiments 4 & 5 directly) suggest that more research is needed to fully understand the relationship between labelling and the context of sorting.

[1] A version of this paper will be revised and resubmitted to Cognitive Science with additional data: Suffill, Branigan & Pickering (2019).

Introduction

Labels not only reflect the way that we categorize objects, but also allow us to communicate and share those categories with others. Labels are thus closely tied to communication. How does the process of labelling affect categories across people, and are these effects tied specifically to coordinative settings (i.e., settings in which people must coordinate their understanding of the world in order to allow for successful communication)? Or does having a coordinative setting simply reinforce the existing effects of labelling on categories? Previous work has shown that using shared labels during categorization can increase the extent to which interlocutors subsequently develop *coherent* categories (i.e., categories that are stable across individuals; Markman & Makin, 1998). Furthermore, even novel labels can have consistent effects upon individual sorters' categories (Lupyan & Casasanto, 2014).

Lupyan (2008) suggested that novel labels help people to select more generalizable dimensions (i.e., dimensions that work well across a range of objects within the category), and that these dimensions tend to be similar across individuals. However, in Markman and Makin's study, participants performed a joint object building task with a partner. Specifically, participants had to use plastic building blocks to build a small structure (i.e., a car or a spaceship). Participants in a pair developed shared labels and used these to build the structure collaboratively. As such, the context was inherently coordinative (i.e., because the coordination of partners' categories and category labels was beneficial to their performance in the building task).

Similarly, in Lupyan and Casasanto's (2014) categorization study, the participants' task involved learning a category and label structure pre-determined by the experimenters, such that the novel labels reflected the dimensions of the stimuli necessary to learn the correct category structure. Hence we do not know whether the greater category coherence between individuals that develops with the use of labels is directly due to the process of labelling, or whether it is specifically an effect of context (i.e., labelling in coordinative contexts, in which a sorter knows that the label must communicate something meaningful about the category it is tied to).

To address this question, we carried out two experiments that investigated the effects of novel labels on category coherence with a coordinative context (i.e., when participants were told that their categories should make sense to another person, and moreover that their categories would be compared to a partner's categories), and without a coordinative context (i.e., when participants were told that their categories should make sense to themselves). If category coherence is a general consequence of using labels, then people who used labels should show greater category coherence than people who did not. Another possibility is that coordinative contexts bolster the effects of labelling, such that − while labels increase category coherence regardless of context − labels increase coherence more so when used within a coordinative setting. But if the effects of labelling are specific to situations in which people must coordinate their categories, then the labelling advantage for category coherence should occur only in coordinative contexts.

1.1 Linguistic categorization and category coherence

People sort objects in similar ways when the categories they form are built upon shared features (i.e., features that are common and sensible to all sorters in a group), such as perceptual or functional information. To an extent, language relies on upon this category coherence across individuals, because people need to have a similar understanding of the world – and how objects within it are labelled – in order to successfully communicate about the world. However, categorization is also affected by language: People categorize objects differently when they use word labels to sort items (*linguistic categorization*), compared with when they sort items without labels (*non-linguistic categorization*).

While linguistic categories tend to be relatively consistent across native speakers of the same language, they vary across speakers of different native languages. For example, Malt et al. (1999) found that when different groups of speakers categorized the same set of stimuli (i.e., 60 container-like objects), native Chinese Mandarin speakers tended to use five different word labels (corresponding to five linguistic categories), whereas American English speakers used seven and Argentinian Spanish speakers used 15. These conventionalized word labels (i.e., existing words with entries in the mental lexicon; Jackendoff, 2002) also have an important impact on category induction: Older children (11-12+ years) and adults rely more on shared labels than perceptual coherence when deciding an object's category membership, such that people prioritize items having shared labels, over perceptual or functional information, during category induction (Sloutsky et al., 2001). As such, people will reliably categorize objects in similar ways given the same labels, under contexts in which the labels are salient to categorization. Therefore, language can increase the coherence of people's categories, when those people speak the same language and, so, use a shared set of linguistic labels for the items they are sorting.

Conventionalized labels influence categorization through top-down effects on the processing of objects' features. For example, hearing that animal X is a 'cat' might mean that we associate with animal X features such as 'has whiskers' and 'purrs', because these are features associated with previously encountered animals also labelled as 'cat' (Gelman, 2003; Gelman & Davidson, 2013). These top-down effects of labelling also extend to early visual processing during categorization: Hearing a redundant label (i.e., a label that provides no new, task-relevant information) still effectively guides visual attention towards target objects. During a visual identification task, people were faster to locate 5s (in a display of 2s & 5s) after hearing 'five', than on trials in which participants did not hear this redundant label, and despite already knowing that their task was to attend to the 5s (Lupyan & Spivey, 2010).

Lupyan (2008) proposed that labels might also affect categorization by causing a shift in how people represent categories, by distorting which object features people successfully store in memory when categorizing and subsequently recalling objects. That is, when speakers apply a category label to an object, it causes the representation of the object to become a mix of its idiosyncratic features and features that are typically associated with the relevant category. Consistent with this claim, participants were worse in a recall task at recognizing objects of furniture that they had previously seen if they had named the object (e.g., they had identified it as a 'table' or 'chair') than if they had made a like/dislike decision for it. This reduction in recall accuracy due to labelling was greatest for the most prototypical objects, presumably because these objects had fewer idiosyncratic features to distinguish them from the category prototype.

Lupyan argued that while this representational shift may reduce recall for some objects, it also supports categorization by helping people to select more generalizable dimensions that work well across a range of objects within the category. Sorters are therefore able to avoid forming categories using too fine and too many dimensions that do not generalize well across a range of objects, and which would lead to a greater and potentially unhelpful number of categories. And when people use fewer, more generalizable dimensions, it becomes more likely that their categories will overlap and be similar to each other.

1.2 Effects of novel labels on category coherence

As well as existing labels with conventionalized meanings, people also frequently encounter novel labels for new or even established objects (e.g., names for new technologies), and such novel labels can also influence category learning in both infants and adults. Waxman and Markow (1995) argued that novel labels actively guide infants' attention to the relevant perceptual similarities across objects, in a way that promotes category learning. Non-word labels (e.g., 'Look at the Timbo!') help guide infants' eye movements to common features across novel objects, so that they increasingly direct their attention more towards shared features than towards dissimilar features – and this leads to enhanced category learning for the novel objects (Althaus & Mareschal, 2014; Althaus & Plunkett, 2016).

Lupyan and Casasanto (2014) argued that, similarly to infants, the presence of novel labels can draw adults' attention to perceptual features across objects, in a way that supports category learning. They had adults categorize novel alien-like stimuli into two pre-determined categories using conventionalized labels (i.e., 'smooth headed' vs. 'pointy headed') or non-word labels (i.e., 'fooves' vs. 'crelches'). Participants performed equally well in learning to label smooth-headed aliens as 'smooth' or 'foove', and pointy-headed aliens as 'pointy' or 'crelch'. In other words, they learnt to assign stimuli to the relevant categories equally well using conventionalized or non-word labels, and more so than in the control condition (in which participants categorized without labels). Although participants were learning pre-determined categories with pre-assigned labels, this finding demonstrates that, under some circumstances, labels that lack explicitly conventionalized meanings within the mental lexicon can still have consistent effects on individuals' categories.

1.3 Importance of coordination for coherence

Shared categories are needed for successful understanding and therefore communication, and people are more likely to have overlapping categories and, so, be more similar to each other if they use fewer, more generalizable dimensions (such as shared, perceptual features that all sorters have experience of; Rosch & Mervis, 1975; Johnson, 1987; Lupyan, 2008). If labels directly cause people to use fewer, more generalizable dimensions for their categories, then labels might increase category coherence between people regardless of coordinative context (i.e., whether the label needs to be able to communicate information about the category to another person, or not).

However, results so far suggesting that labels can increase category coherence between people come from studies involving an explicitly coordinative context (e.g., Markman & Makin, 1998; Suffill, Branigan & Pickering, 2016). Therefore, the effects of labelling in these studies might have been due to the context in which labelling occurred (i.e., a context in which the need to coordinate categories with a partner was foregrounded). The labelling effects found in Lupyan and Casasanto (2014) might also have been affected by a coordinative context, since participants knew that the non-word labels communicated important information regarding the pre-determined categories that they had to learn for the task.

How, then, might coordinative contexts affect the way that people use labels to categorize? One possibility is that labels directly reorganize how people select dimensions for categorization, in a way that makes them categorize more similarly to each other (i.e., labelling directly increases category coherence). In that case, people would tend to categorize in similar ways to each other whenever they used labels, compared to when they did not use labels, and irrespective of the context. But in this case, the need for coordination with a partner might simply reinforce the effects of labelling, such that labelling increases category coherence both with and without a coordinative context, but this increase in coherence is greater when labelling occurs with a coordinative context, compared to without a coordinative context.

Alternatively, labels might affect how people select dimensions for categorization in a way that is tied specifically to the communication of those categories. That is, labels might act as a device for coordinating categories with others only in contexts that explicitly require coordination (Clark, 1996). In that case, novel labels might increase category coherence specifically in contexts that involve coordination (coordinative contexts). For example, labels might influence people to select more generalizable category dimensions when they are told that their categories must make sense to others, compared to (a) when they sort without labels, or (b) when they sort with labels but without their categories needing to make sense to others (non-coordinative contexts). In addition, as categories that would make sense to one partner are likely to also make sense to other people (see Garrod & Doherty, 1994), using labels in a coordinative context might lead people to develop greater category coherence not simply with one partner, but also with a community of people undertaking the same task of sorting. Under this account, labels might increase category coherence between people who sort in a coordinative context even if they are not directly exposed to each other's categories.

1.4 Current study

We have seen that novel labels – like conventionalized labels – can have consistent effects on the way that people form categories, by affecting which object dimensions they select for categorization (Lupyan et al., 2007; Lupyan & Casasanto, 2014; Suffill et al., 2016). However, it is not clear whether the effects of novel labels on category coherence might occur in any context, or whether they might instead be specific to contexts in which there is a need for people to coordinate their categories with others. To test between these possibilities, we carried out two online experiments, in which we investigated the effects of novel, non-word labels on category formation and category coherence for real world items (i.e., grayscale images of mountains), when participants did or did not have the context of forming categories that would make sense to another person.

Participants sorted items into two self-determined groups across three rounds. Participants sorted items into groups either by using randomly-generated, non-word CVC labels (which were different for every participant in that condition), or without using any labels for sorting. We also manipulated the context in which participants sorted items: in Experiment 4, participants were instructed to sort groups of items in a way that made sense to them alone; in Experiment 5, participants were instructed that they would be paired (but not interact) with a partner, to whose categories their own categories would be compared, and so that their categories should make sense to another person.[2]

We predicted that novel labels might change the way that people sorted by affecting which dimensions they selected for category formation, and that using labels might therefore lead to people sorting the items more similarly to each another, by causing them to select more generalizable dimensions based on shared, perceptual features. Hence we predicted that overall, there would be greater category coherence between people who sorted using labels versus those who sorted without using labels.

However, we also predicted that the effect of novel labels on category coherence might be affected by the context of sorting. That is, labels might affect the selection of category dimensions in a way that ties specifically to communicating those categories to others, and so novel labels might primarily increase category coherence in contexts that involve the coordination of categories. If so, then participants who sorted with labels within a coordinative context would have greater category coherence, than participants who sorted with labels but without a coordinative context. Hence, we predicted that category coherence would be greater for participants who sorted with labels in Experiment 5 (coordinative context) compared to participants who sorted with labels in Experiment 4 (non-coordinative context). Moreover, if labelling effects depend entirely on a coordinative context, then we would expect to find greater category coherence between people who sorted using labels versus those who sorted without using labels only in Experiment 5; and there would be no such effects in Experiment 4.

Experiment 4

2.1 Methods

2.1.1 Participants. Participants were 100 native monolingual English speakers (61 female) from the U.S. and the UK. Ages ranged from 18-35 years (*M* = 28.04, *SD* = 4.81).

[2] In fact, as we discuss below, our analyses compared all participants within a condition to each other (i.e., we did not compare individual pairs).

Participants were recruited through a participant hosting website, Prolific Academic (https://www.prolific.ac.uk). The University of Edinburgh's Ethics Committee approved this study.

2.1.2 Stimuli. Stimuli comprised 72 greyscale images of mountains from a range of locations (see Fig. 1 for a subset). We chose images of mountains because participants were unlikely to have strong preconceptions about how groups of mountains should be categorized. We created three sets (A-C, see Appendices J-L), each comprising 24 items (i.e., images). Participants sorted one set in each round of the experiment. We selected items that presented a relatively small number of dimensions by which to vary (e.g., height, sharpness, number of peaks, snow coverage, etc.). Additionally, we randomly generated 50 unique pairs of novel, non-word labels each comprising a CVC structure (e.g., 'rah', 'jib'; see Appendix M for full set). Labels remained the same for participants across rounds.

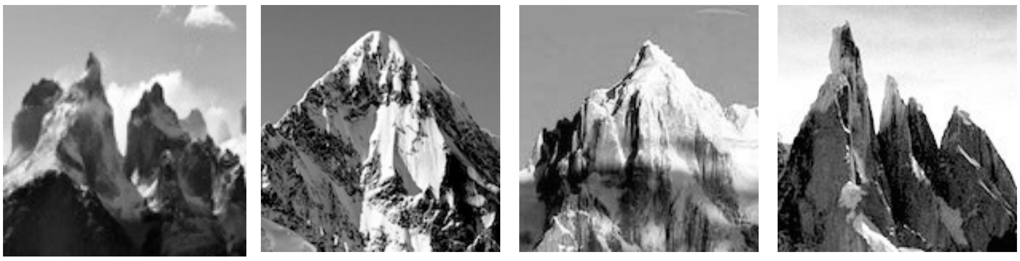


Fig. 1. Subset of stimuli (from Set A).

2.1.3 Design. There was one factor: Labels (between-participants; with-labels vs. no-labels). We randomly assigned 50 individuals to each of the two conditions. In the with-labels condition, participants sorted stimuli into two groups using novel, non-word labels. In the no-labels condition, participants sorted stimuli into two groups without using labels. The dependent variable was group coherence, which was the extent to which participants' categories were coherent with those of other participants in the same condition (i.e., the extent to which they put the same items in the same categories).

2.1.4 Procedure. Participants were recruited via Prolific Academic, and the experiment was run via the website Qualtrics (https://www.prolific.ac.uk). Once participants had confirmed their consent for participation, they were re-directed to the experiment page, in which they began the first round of sorting. Instructions told participants that they would see a group of greyscale images depicting mountains (see Fig. 2A). These images appeared on the left-hand side of the screen and were individually randomized for each participant. They were asked to sort these images into two categories (i.e., 'Please sort them in a way that makes sense to you.') by dragging the items into one of two predefined spaces onscreen (see Fig. 2B). Participants could drag items across to these spaces in any order and could change their groupings freely during a round.

In the with-labels condition, these spaces were labelled with a pair of non-word labels that were unique for participant (also, these labels remained the same across rounds for each participant). In the no-labels condition, the two predefined spaces were unlabelled. Participants were told that they could have any number of items in each group, as long as neither group contained fewer than 8 items when they finished. They were also instructed that they must sort every item (i.e., they could not leave any items uncategorized). Instructions were repeated in each round. Participants sorted set A in round 1, set B in round 2, and set C in round 3 (set order was fixed to allow item comparisons across all participants within each condition). Participants could not return to a round once they had proceeded to the next and had a total of 15 minutes to complete the task.

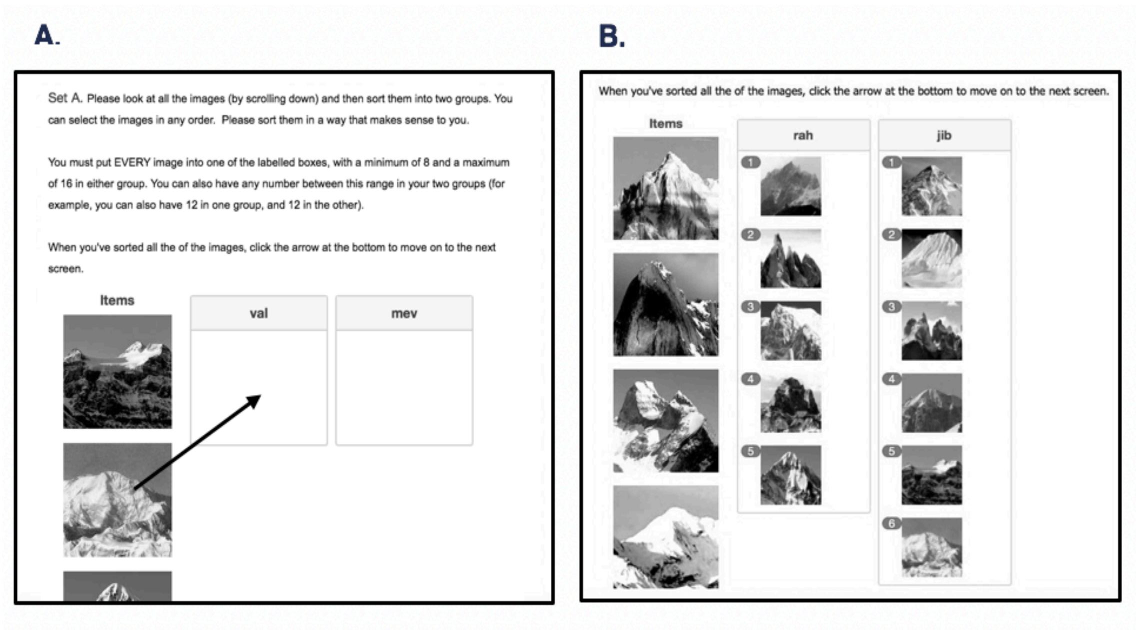


Fig. 2. Screenshots of how a participant might begin placing items into a category (A); and how items looked once they had been placed into the categories (B) (with-labels condition).

2.2 Results

2.2.1 Analysis. We investigated the effects of labels at the group-level by comparing category coherence between every possible pairing of participants within each condition within Experiment 4. That is, we took every participant from the with-labels condition and compared their categories to every other participant from the with-labels condition; similarly, we took every participant from the no-labels condition and compared their categories to every other participant from the no-labels condition. We included Round (1-3) when modelling random effects, but did not include it as a fixed effect (i.e., because we always used Set A, B and C in Round 1, 2 and 3 respectively, to allow item comparisons across all participants within each condition). We predicted that the use of novel labels during categorization might directly influence which dimensions people selected for sorting, and that these dimensions might be coherent across people. If these predictions held, then category coherence would be higher even in a non-coordinative context for the group of participants who sorted with labels, compared to the group of participants who sorted without labels. However, if the effects of labelling on category coherence were dependent on having a coordinative context, then the with-labels condition would not show greater coherence than the no-labels condition.

Group category coherence. In order to compare category coherence across every possible pair, we implemented the CCM (Romney et al., 1986; Malt et al., 1999; Ameel et al., 2005; White et al., 2016). For every participant, we coded whether they put each possible pair of items (24x23/2 = 276 item pairs per round) into the same category or not. If a participant placed two items into the same category, that item pair would be coded as 1; if not, the item pair was coded as 0. We then used this data to compute a measure of association between each possible pairing of participants within a condition by calculating a proportional score between 1 and 0. We used these scores in our analyses. For example, if a pairing of participants matched on all 276 item pairs within a round, they would be given a score of 1. If they matched on no item pairs, they would be given a score of 0. We calculated proportional scores for every possible pairing of participants for the with-labels condition and the no-labels condition, and across rounds 1-3 (although Round was used only as a random effect in the analysis). We then averaged these scores to create an average proportion of category coherence for each individual in the study.

Descriptive Statistics. Average CCM scores (SD) across all possible pairs in the with-labels and no-labels conditions are summarized in Table 1. In addition, individual data points (average by participant) are plotted across the conditions as bee-swarm plots (see Figures 3 & 4).

Table 1.

Experiment 4: Average CCM scores (SD) across rounds for the Labels conditions.

Labels	Round	Mean	SD
No labels	1	0.292	0.047
	2	0.271	0.029
	3	0.265	0.028
With labels	1	0.268	0.028
	2	0.269	0.024
	3	0.298	0.069

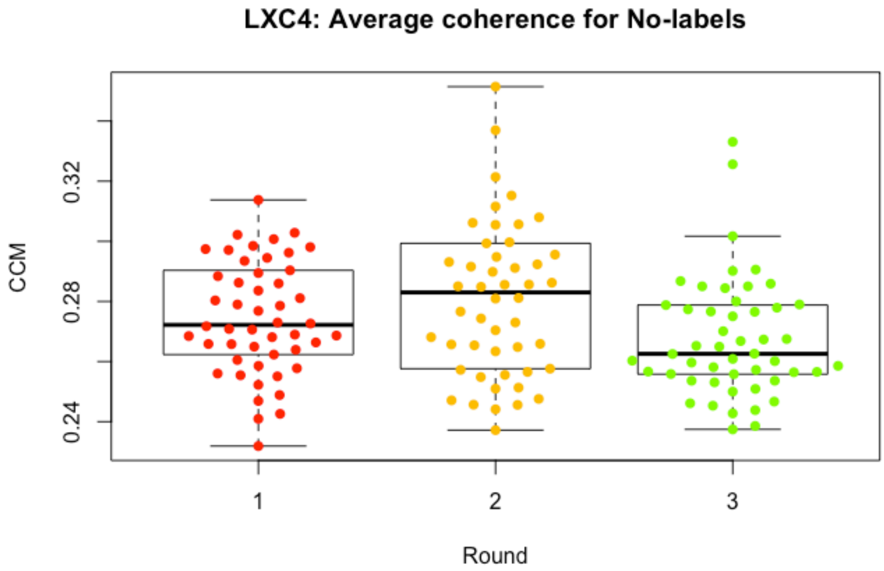


Fig. 3. Experiment 4; Average category coherence (CCM) by Round (1-3) for participants in the No-labels condition.

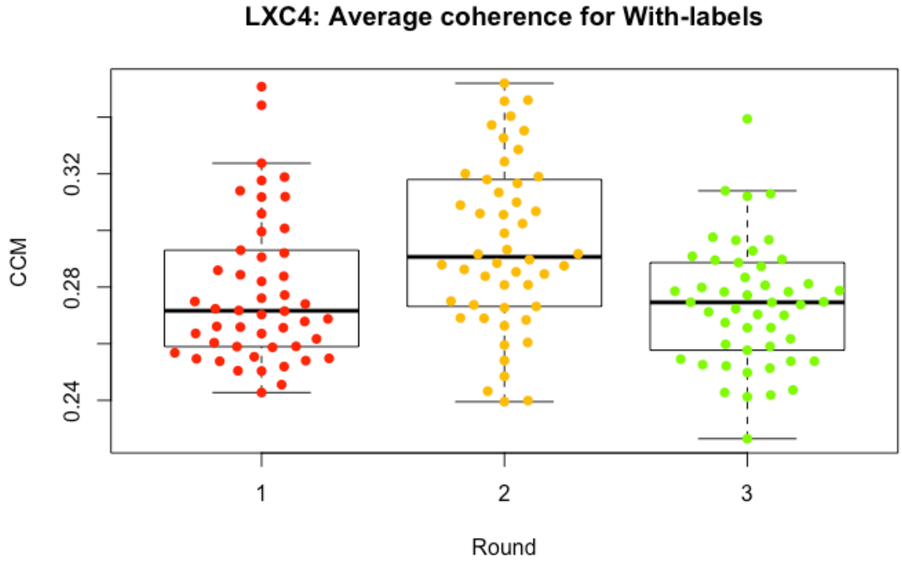


Fig. 4. Experiment 4; Average category coherence (CCM) by Round (1-3) for participants in the With-labels condition.

LME analysis. In order to normalize the sampling distribution of the proportional CCM scores, the scores were Z-transformed with $Z=0.5*\ln[(1+r)/(1-r)]$. To test for the effects of Labels on category coherence, data were analyzed in R 3.2.1 (R Core Team, 2015), with the lme4 package, version 1.1-8 (Bates et al., 2014). This approach allowed us to account for random variance due to differences between participant pairings and round. The threshold for statistical significance was set at $|t| > 2$ (Baayen, 2008). A backwards, stepwise elimination approach was used to select factors for the final model, using likelihood ratio tests to compare models. Models included random slopes for the factor Labels by Person and by Round (1-3). The model included Labels (with-labels vs. no-labels) as a fixed effect, with the reference level for Labels set as no-labels. This model was a significantly better fit than the null model ($X^2(1) = 18.29$, $p < .001$). Under this model, there was not a significant effect of Labels on category coherence ($t < 2.00$) (see Tables 2 & 3).

Table 2.

Experiment 4: Beta, standard errors and t-values for fixed effects on category coherence Z-score. Model fit by REML.

Fixed effects	***β***	**S.E.**	***t***
Intercept	0.00	0.06	0.00
Labels	0.03	0.20	0.17

Table 3.

Experiment 4: Variance and residual for random effects. Model fit by REML.

Random Effects		
Person	Labels	0.16
Round	Labels	0.11
Residual		0.78

No. of Observations = 300.

2.3 Summary of Experiment 4

Contrary to previous findings that novel labels can increase category coherence across people, we found no effect of sorting with labels on the similarity of people's categories. However, this experiment involved a non-coordinative context, in the sense that there was no requirement for participants to sort in a way that would make sense to another person. To test whether the effects of labels on category coherence are specific to contexts in which there is a need to coordinate categories, in Experiment 5 we again manipulated whether participants sorted with or without novel labels, but this time in a context in which the task instructions foregrounded the need for categories to make sense to a partner.

Experiment 5

Experiment 5 had the same design as Experiment 4, but presented a context that emphasised coordination with others: All participants were told that they should sort the stimuli in a way that would make sense not only to themselves, but also another person, and moreover that their categories would be compared to those of another person who had sorted the same stimuli, and – in the with-labels condition – used the same non-word labels to categorize (although, as in Experiment 4, each participant in the with-labels condition sorted with a different pair of non-word labels).

3.1 Methods

3.1.1 Participants. Participants were a further 100 native monolingual English speakers (57 female) from the U.S. and the UK, who did not take part in Experiment 4. Ages ranged from 18-35 years (M = 27.84, SD = 4.90).

3.1.2 Stimuli. Stimuli (i.e., the images of mountains and non-word labels) were identical to Experiment 4.

3.1.3 Design. The design of the experiment was identical to Experiment 4 (i.e., one independent factor of Labels). We randomly assigned 50 individuals to each of the two conditions. As in Experiment 4, the dependent variable was group category coherence within each of the Labels conditions.

3.1.4 Procedure. The procedure was identical to Experiment 4, except for the amendments to the instructions to participants in order to produce a coordinative context. As such, participants in the no-labels condition were given the additional instructions regarding the items: 'Please sort them in a way that makes sense to you, but that would also make sense to another person' and 'You'll be assigned a partner who will separately sort the same items – we'll then compare the way you sorted the items to how they sorted them'. In contrast, participants in the with-labels condition were given the additional instructions: 'Please sort them in a way that makes sense to you, but that would also make sense to another person' and 'You'll be assigned a partner who will separately sort the same items – we'll then compare the way you sorted the items using the given labels, to how they sorted them'.

3.2 Results

3.2.1 Analysis. As in Experiment 4, we investigated the effect of Labels at the group-level by comparing category coherence between every possible pair of participants within a condition. We again hypothesized that the use of novel labels during categorization might influence which dimensions people selected for sorting, and that these dimensions might be coherent across people, but − given that labels did not increase category coherence in a non-coordinative context (Experiment 4) – we additionally hypothesized that the effects of labels might require a coordinative context. If these predictions held, then category coherence would be higher for the group of participants who sorted with novel labels, compared to the group of participants who sorted without labels, within the coordinative context of Experiment 5.

Group category coherence. We again implemented the CCM to investigate the effects of labelling at the group-level by comparing category coherence between every possible pairing of participants within each condition. We calculated proportional scores for every possible pairing from the with-labels condition and every possible pairing from the no-labels condition. We again averaged these scores to create an average proportion of category coherence for each individual in the study. As before, we included Round (1-3) when modelling random effects, but did not include it as a fixed effect.

Descriptive Statistics. Average CCM scores (SD) across all possible pairs in the with-labels and no-labels conditions are summarized in Table 4. In addition, individual data points (average by participant) are plotted across the conditions as bee-swarm plots (see Figures 5 & 6).

Table 4.

Experiment 5: Average CCM scores (SD) across the two Labels conditions.

Labels	Round	Mean	SD
No labels	1	0.277	0.073
	2	0.284	0.031
	3	0.267	0.020
With labels	1	0.281	0.031
	2	0.296	0.032
	3	0.276	0.027

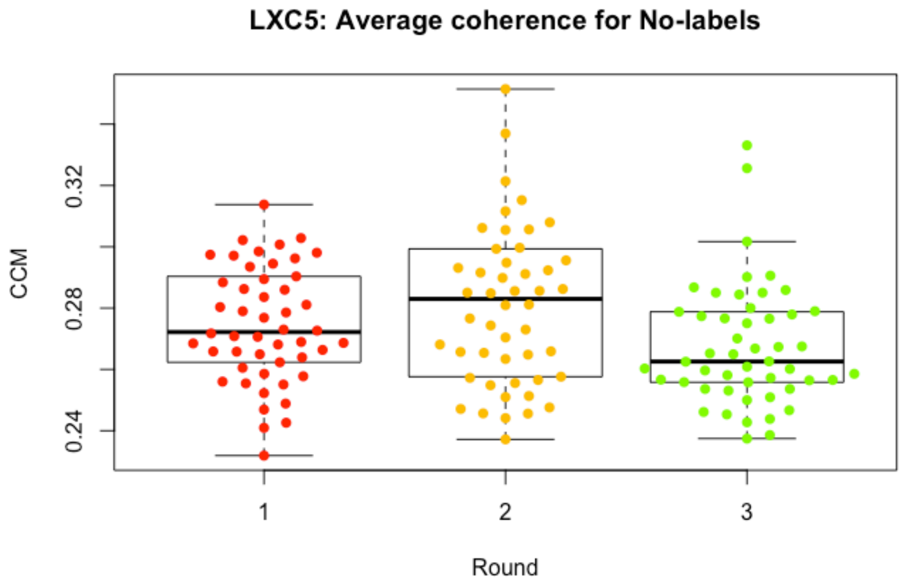


Fig. 5. Experiment 5; Average category coherence (CCM) by Round (1-3) for participants in the No-labels condition.

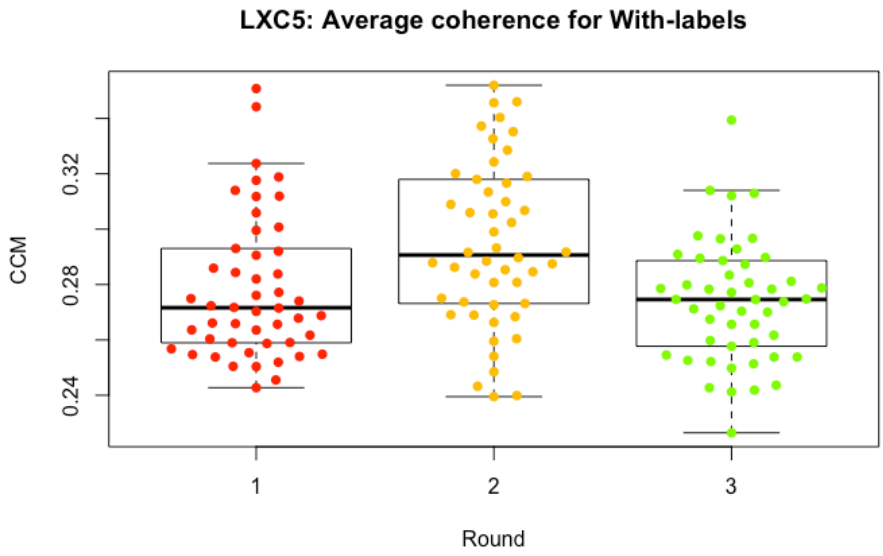


Fig. 6. Experiment 5; Average category coherence (CCM) by Round (1-3) for participants in the With-labels condition.

LME analysis. Data were again normalized to Z-scores. To test for the effects of Labels on category coherence, data using an LME analysis. Again, a backwards, stepwise elimination approach was used to select factors for the final model, using likelihood ratio tests to compare models, and models included random slopes for the factor Labels by Person and by Round. The model included Labels (with-labels vs. no-labels) as a fixed effect, with the reference level for Labels set as no-labels. This model was a significantly better fit than the null model ($X^2(1) = 3.84$, $p < .05$). Under this model, there was a significant effect of Labels, such that participants who sorted with labels had greater category coherence than those who sorted without labels (β= 0.27, SE = 0.13, t = 2.06) (see Tables 5 & 6).

Table 5.

Experiment 5: Beta, standard errors and t-values for fixed effects on category coherence Z-score. Model fit by REML.

Fixed effects	*β*	S.E.	*t*
Intercept	0.00	0.06	0.00
Labels	0.27	0.13	2.06

Table 6.

Experiment 5: Variance and residual for random effects. Model fit by REML.

Random Effects		
Pair	Labels	0.70
Round	Labels	0.00
Residual		0.82

No. of Observations = 300.

3.3 Summary of Experiment 5

In Experiment 5, we again manipulated whether participants sorted with or without novel labels, but this time in a context in which the task instructions stressed the need to coordinate with a partner. In contrast to Experiment 4, Experiment 5 showed greater category coherence for people who sorted with labels compared to those who sorted without labels within a coordinative context. In order to directly compare these contexts of sorting, we next conducted a combined analysis of the data.

Combined Analysis

In order to directly compare the effects of context and labels on category coherence, we conducted an analysis on the combined data from Experiments 4 and 5.

4.1 Results

4.1.1 Analysis. Again, we investigated the effect of Labels at the group-level by comparing category coherence between every possible pairing of participants within the with-labels condition and every possible pairing of participants within the no-labels condition. But in addition, we included Context (coordinative [Experiment 4] vs. non-coordinative [Experiment 5]) as a predictor.

Group category coherence. We again calculated proportional scores for every possible pairing of participants from the with-labels condition and every possible pairing of participants from the no-labels condition, in Experiment 4 and in Experiment 5 separately. As before we included Round (1-3) when modelling random effects, but did not include it as a fixed effect.

LME analysis. Data was normalized to Z-scores and analyzed using an LME analysis. Again, a backwards, stepwise elimination approach was used to select factors for the final model, using likelihood ratio tests to compare models. Models included random slopes and intercepts for the factor Labels by Person and by Round. The model included Labels (with-labels vs. no-labels) and Context (non-coordinative vs. coordinative) as fixed effects, with the reference level set as no-labels:non-coordinative. This model was a significantly better fit than the null model ($X^2(7) = 23.68$, $p < .001$). Removing the interaction term did not significantly reduced model fit ($p > .05$). Removing either of the main effects did not significantly reduce fit ($p > .05$), but a model with only Context as a fixed effect was a significantly better fit, than a model with only Labels as a fixed effect ($X^2(0) = 0.66$, $p < .001$). Therefore, the model of best fit included only Context as fixed and random effects. Under this model, there was not a significant effect of Context (see Tables 7 & 8).

Table 7.

Experiments 4 & 5, combined analysis: Beta, standard errors and t-values for fixed effects on category coherence Z-score. Model fit by REML.

Fixed effects	**β**	**S.E.**	**t**
Intercept	-0.00	0.04	0.00
Context	0.04	0.01	0.35

Table 8.

Experiments 4 & 5, combined analysis: Variance and residual for random effects. Model fit by REML.

Random Effects		
Person	Intercept	0.12
	Context	0.04
Round	Intercept	0.00
	Context	0.04
Residual		0.82

No. of Observations = 600.

General Discussion

In two experiments, participants individually sorted grayscale images of mountains into two groups. Participants either sorted items into groups without labels, or sorted items into groups using two novel, non-word labels (which were different for each participant within that condition). If using labels led people to develop more similar categories and, so to have greater category coherence, then people who used labels should show greater category coherence than people who did not use labels, and this greater coherence should occur even in the absence of a coordinative context (as in Experiment 4). But if the labelling effect is specific to situations in which people must coordinate (e.g., in order to have successful communication), then it should only occur in a coordinative context (as in Experiment 5). Individual analyses of the experiments suggested that: (1) in a coordinative context (i.e., a context in which the labels should communicate something about the categories to another person), people who sorted with novel labels showed increased category coherence as a group compared to people who sorted without novel labels (Experiment 5), but not in a non-coordinative context (Experiment 4).

Specifically, the results from Experiment 4 demonstrated no difference in category coherence for people who sorted with or without labels. However, results from Experiment 5 showed greater category coherence among people who sorted with labels, than those who sorted without labels, in a coordinative context. However, the null results found in the combined analysis make these effects harder to interpret since they do not allow us to directly compare the effects of labels across the two experiments. As such, results will be tentatively discussed in terms of the individual experiments.

5.1 Novel labels in a non-coordinative context

In Experiment 4, participants who sorted with novel, non-word labels displayed no difference in their category coherence compared to those who sorted without labels. In the absence of a coordinative context and shared labels (i.e., labels were not shared because each participant in the with-labels condition sorted using unique labels), labelling may have led to greater variation in the sort of distinctions people chose when categorizing items, leading to no benefit in coherence for participants who sorted without labels.

Other research has shown that novel labels do not always lead to increased category coherence among people, and indeed that novel labels can sometimes reduce coherence in certain settings that ask us to take into account other people's ways of categorizing. Silvey (2014) failed to find beneficial effects of labelling upon category coherence between communicating pairs who were negotiating category structure (when interactions were computer-mediated and did not involve interacting with a partner directly). As such, when two individuals are sorting the same items with different novel labels, the lack of need for coordination may reduce the beneficial effects of labelling in categorization.

5.2 Novel labels in a coordinative context

Conversely, we suggest that in Experiment 5 the coordinative nature of the context caused the label to become a potential means for communication, and − through this – a focus for the coordination of people's categories (Clark, 1996). Lupyan (2008) argued that labels affect people's categories by influencing them to select more generalizable dimensions that work well across a range of objects, in comparison to when people categorize items without labels. We propose that in doing so, labels can also increase the coherence of people's categories by influencing people to select perceptually shared dimensions.

However, for novel labels with no conventionalized meaning, the context of categorization also plays an important role: when the context is coordinative, labels influence the sorter not only to select certain dimensions for categorization, but moreover to select those dimensions that would be sensible to another person. As such, the label must be applied to the referents within each category in a way that would make sense to another person doing the same task, and as if task partners had to use those labels to successfully communicate with each other about objects from the categories (Clark & Brennan, 1991; Clark, 1996).

5.3 Coordination without interaction

However, unlike previous research examining how people coordinate labels for objects and therefore linguistic categories (e.g., Garrod & Doherty, 1994; Clark & Brennan, 1991; Markman & Makin, 1994; Suffill et al., 2016), our participants were never exposed to or interacted with each other, and each sorter within the with-labels condition used a unique pair of non-word labels. Despite this, people's categories still had greater coherence when they used non-conventionalized, novel labels during sorting within a coordinative context, compared with people who sorted without labels within a coordinative context. Given that participants in each condition had no opportunity to interact with each other or provide feedback to each other, this finding suggests that the category dimensions people selected when using labels for coordination were most likely based on shared, perceptual features of the stimuli, and not on any particular aspects of the labels themselves (e.g., word form features).

We suggest that there are fundamental commonalities regarding which dimensions people find easy to abstract, in the same way that there are fundamental commonalities regarding the way in which people perceptually experience objects (Rosch & Mervis, 1975; Johnson, 1987). In other words, the novel labels led to greater category coherence in the coordinative context because they helped people to identify appropriate category dimensions when the sorter considered another person's perspective – and what was appropriate for one other person (i.e., the supposed partner that participants were asked to consider when sorting) was likely to be appropriate for many other people within that condition.

Lastly, our combined analysis of the data allowed us to assess whether there was a general effect of context on category coherence (i.e., whether being in a coordinative or non-coordinative context affected coherence, regardless of access to novel labels). Within this analysis, there was no surprisingly no effect of context, which makes our overall results harder to interpret across both experiments. Despite this, the results of Experiment 5 suggest that labels can sometimes be beneficial to the formation of greater category coherence across sorters. This reiterates the importance of language, as a tool for coordination in communicating with others (Clark, 1991).

5.4 Conclusion

Similarly to conventionalized labels, novel non-word labels influence which object dimensions people select for categorization and, by doing so, can increase the coherence of people's categories. However, we show that the context in which people sort can also be crucial to the effect of novel labels on people's category coherence. In contexts in which the goal is non-coordinative (and therefore individualistic), novel labels may not benefit category coherence. In coordinative contexts, the same novel labels (that vary across individuals) can lead to increased category coherence, even without exposure or feedback between people. The results of Experiments 4 and 5, therefore, provide further evidence that novel labels serve to direct people's attention to more generalizable object dimensions that people tend to agree on, therefore, increasing the potential for people's categories to overlap. However, the null results of our combined analysis suggest that more research is needed to fully understand the relationship between labelling and the context of sorting.

Chapter 5

Effects of interaction and coordination on category coherence in L1-L2 dialogues

Interlocutors discussing objects tend to become more coherent in how they label and therefore linguistically categorize those objects. Therefore, people's linguistic categories can be influenced by interaction. But second language (*L2*) speakers' linguistic categories differ from native (*L1*) speakers' categories, because an L2 speaker's category structure is influenced by the structure of their L1. When two speakers start with significantly different linguistic categories, does interaction lead to the greater coherence of their categories? And if so, what conditions are required to increase coherence between L1 and L2 speakers? We investigated (a) whether discussion increases the similarity of people's categories (*category coherence*) in L1-L2 pairs, (b) how the need for coordination between partners affects this process, and (c) whether these effects lead to differences in category change across L1 versus L2 speakers. L1-L2 pairs individually categorized dishware with intermittent interaction: in Experiment 6, participants discussed categories, or unrelated images; in Experiment 7, all participants discussed their categories, but some pairs did so with a coordinative goal of having more similar categories to each other. Discussion of categories did not increase the coherence of pairs' categories (*pair coherence*), and the addition of coordinative discussion did not change this. Additionally, we examined coherence across all participants within each condition (*group coherence*). Discussion of categories differentially affected the similarity of L1 speakers' categories to those of other L1 speakers, but did not affect the similarity of L2 speakers with other L2 speakers, or the similarity of L1 and L2 speakers to each other (Experiment 6). The addition of a coordinative goal with discussion of categories did not increase group coherence, over discussion of categories with a non-coordinative goal (Experiment 7). Results suggest that the effects of category-relevant discussion on category structure and coherence are affected by the status of the speaker, on the basis of whether they are an L1 or L2 speaker of the language. Secondly, they show that explicit coordination does not always lead to increased category coherence between pairs in L1-L2 dialogues.

Introduction

Discussion about and exposure to a person's categories play a fundamental role in learning how that person conceives and labels particular concepts, and hence in developing categories that are similar. Category structure can be stable across people, when the categories are built upon shared features that all sorters have access to, like perceptual or functional information about items. To an extent, language relies upon category coherence across people because people need to have a similar understanding of the world – and how objects within the world are labelled – in order to communicate successfully. However, different languages can carve up the world in different ways, leading to different linguistic categories for objects (i.e., differences in how objects are grouped together under linguistic labels).

Category coherence can be partly due to interaction – people interact with each other and as a result converge on their linguistic categories. It is possible that such convergence through interaction is limited to speakers of the same language who already have comparatively similar linguistic categories. But alternatively, speakers of different native languages – that have very different linguistic categories – might also converge. And if so, it may be that such convergence is a consequence of interaction. Secondly, it may be that the interlocutors – especially interlocutors with very different linguistic categories – need to have a goal of achieving such convergence within interaction, in order to converge upon common patterns for linguistically categorizing items.

To address these questions, we used an interactive paradigm with L1-L2 speaker pairs (specifically, pairs comprising a L1-English and a L1-Mandarin/L2-English speaker) that had very different linguistic categories for dishware, to examine whether discussion between speakers of different native languages led to greater coherence in their categories (*pair coherence*). Secondly, we examined how having a non-coordinative versus having a coordinative goal during discussion might affect the process of categorization, and the coherence of pairs' categories. Furthermore, we asked whether discussion of categories with a specific partner would also affect category coherence across people as a wider group, by comparing the coherence of all possible pairs within each condition, for the groups of L1-L1, L2-L2 and L1-L2 speakers (*group coherence*). This group coherence reflected whether sorters became more similar to an average way of sorting as a group (i.e., through having exposure to an L1/L2 partner, or not).

1.1 Factors affecting categorization

Several factors affect how people categorize objects, including the perceptual features of objects. Non-linguistic, perceptual categories (i.e., categories based upon the perceptual features of items) for real world objects – such as food containers and dishware – tend not to vary significantly across people (Malt et al., 1999; Laskowski, 2010). This is because perceptual categories are based upon physical aspects such as size, shape, and colour. Thus, because people share perceptual experience of the same structured world, their perceptual categories tend to overlap (Rosch & Mervis, 1975; Johnson, 1987).

However, categorization is also affected by language. People categorize items differently when they use word labels than when they do not use word labels to sort objects. And speakers of different native languages categorize objects similarly when they do not use labels, but categorize more differently to each other when they use conventionalized linguistic labels for their categories. For example, when categorizing the same set of stimuli (i.e., 60 container-like objects), a group of native Chinese Mandarin speakers tended to use five different word labels, whereas American English speakers used seven and Argentinian Spanish speakers used 15 (Malt et al., 1999). People may therefore vary in how they categorize objects on the basis of language (i.e., because they speak different languages and therefore rely upon different sets of linguistic labels for categorization).

1.2 Categories in the L2

Given that linguistic categories differ across speakers of different languages, bilinguals may have knowledge of two languages that map words onto referents in different ways (Ameel et al., 2005; Ameel, Malt, Storms & Van Assche, 2009). However, it is also possible that a bilingual's linguistic categories might influence each other across the two languages. Ameel et al. (2005, 2009) examined word-to-referent mappings in Dutch–French bilinguals by comparing the way that they named objects in their two languages. They found that when Dutch-French bilinguals categorized real world objects (i.e., dishware & bottles) their naming patterns in both languages converged upon a common naming pattern. This convergence also led to changes in the structure of bilingual linguistic categories, such that bilingual categories differ from monolingual categories (i.e., the linguistic categories in both of a bilingual's languages).

Additionally, in late bilinguals the development of linguistic categories in the second language is heavily influenced by knowledge of the first. This, again, means that the linguistic categories and naming patterns of the bilingual's L2 language significantly differ from the sorting patterns of an L1 speaker (Malt, Li, Pavlenko, Zhu & Ameel, 2015). As such, results suggest that when bilinguals (i.e., with either early or late exposure) and monolinguals interact, their concepts and categories are not necessarily coordinated, and to a lesser extent than we might expect between two native, monolingual speakers of the same language (Costa et al., 2008).

1.3 Coherence in L1-L1 and L1-L2 dialogues

Between native speakers of the same language, interaction between interlocutors can lead to alignment on several levels of linguistic and conceptual representation, such that interlocutors coordinate their mental states (Pickering & Garrod, 2004). Therefore, having exposure to another person's categories can be crucial in learning how that person conceives and labels certain concepts, such that greater discussion and exposure leads to the development of more similar linguistic categories between interlocutors. For example, using the same labels to refer to items during an interactive, collaborative task can subsequently lead to individuals separately categorizing those items more similarly to one another, than if they did not use shared labels during the interaction (Markman & Makin, 1998).

However, accounts such as Pickering and Garrod's (2004) *interactive alignment,* focus on fully competent, native speakers of the same language. In dialogue between native monolingual speakers of the same language – although naming patterns for objects may not be identical – an extensive overlap already exists between speakers. In L1-L1 dialogues, interlocutors rarely need to explicitly negotiate their coordination of linguistic categories (i.e., using the same label for an object) (Garrod & Anderson, 1987). Instead, this coordination of linguistic categories is argued to occur on a predominantly automatic and implicit basis between L1 speakers of the same language.

In contrast, the substantial differences in the linguistic categories of L1 and L2 speakers leads to less overlap in the naming patterns of L1 and L2 interlocutors, compared to the overlap in naming patterns between two L1 speakers. This means that the development of category coherence through interaction is potentially attenuated in L1-L2 dialogues, compared with cases in which interlocutors have more similar categories (i.e., because they speak the same native language and, therefore, have very similar label-to-referent mappings).

Despite this, there is evidence that L1s and L2s can coordinate on several levels of linguistic representation (e.g., lexical and syntactic alignment; Schoonbaert et al., 2007). Costa et al. (2008) theorized that L1 and L2 interlocutors may coordinate their representations through less automatic or implicit routes than would be expected between two L1 speakers. For example, L2s could coordinate with the choices (e.g. lexical, syntactic and conceptual) used by an L1 speaker, in order to improve their language acquisition. That is, the L2 speaker recognizes the L1 speaker as a more competent speaker of the language, and therefore attempts to use choices put forth by the L1 speaker. As such, the L2 speaker aligns with the L1 speaker in order to become more L1-like in, for example, the way they might label objects. In this way, it also allows the L2 speaker to test their usage of a more L1-like choice against the reaction of the L1 speaker (Mackey et al., 2000).

L1 speakers could also shift their categories towards those of L2 speakers for the purpose of accommodation. For example, L1 speakers may try to explicitly accommodate L2 speakers by changing their linguistic choices towards simplified speech that is more easily understood by the L2 speaker (Arthur et al., 1980). As such, there are several mechanisms by which the linguistic categories of L1 and L2 speakers might still become more coordinated through interaction in L1-L2 dialogues, despite the substantial differences in their linguistic categories. But for these less automatic mechanisms to occur and, thus, increase L1-L2 category coherence – the context of the interaction might require a more explicit goal of coordinating categories.

A final point we considered is whether the interaction between one L1-L2 pair has effects on the category coherence between L1s and L2s as a wider group. Specifically, when an L2 speaker interacts with an L1 partner about their categories, does this yield any benefits for category coherence between that L2 speaker and other L1 speakers whose categories they have not been exposed to (and vice-versa for an L1 speaker with other L2 speakers)? If interaction with one partner leads to L1 and L2 speakers having increased category coherence, then this might be due to them selecting strategies that make sense to both partners. And what is suitable for one partner is likely to be suitable across many people, since the need to coordinate with one person could result in the sorter relying upon shared conventions for sorting (Garrod & Doherty, 1994). Group coherence, therefore, reflects whether sorters became more similar to an average way of sorting, through having some form of interaction with a partner.

1.4 Current study

We were interested in establishing the effects of interaction on category coherence before and after interaction. Specifically, we examined whether interaction can lead to increased category coherence between speakers with significantly different linguistic categories. Secondly, we examined how having coordinative versus non-coordinative goals affected this process. Furthermore, we asked whether the interaction between one L1-L2 pair had effects on the category coherence of L1s and L2s as a wider group, by comparing all possible L1-L1, L2-L2 and L1-L2 pairs within each experimental condition.

In two experiments, 80 pairs comprising an L1-English and an L2-English/L1-Mandarin speaker individually categorized dishware as ‘bowl’ or ‘plate’ across eight rounds, and interacted with their partner at intervals between rounds in the *interaction phase* (i.e., rounds 3-6). The critical rounds for measuring category coherence were rounds two (*pre-test*) and seven (*post-test*), since they reflected a sorter’s categories before and after interaction with a partner, and required participants to use linguistic labels (i.e., ‘bowl’ vs. ‘plate’) for their categories.

In Experiment 6, we manipulated whether or not participants interacted about and were exposed to their partner’s categories at the intervals between sorting, in the interaction phase. By doing so, we aimed to investigate whether discussion about a partner’s categories would increase category coherence between L1s and L2s, in comparison to discussion about something irrelevant to the categories (i.e., discussion of unrelated images of landscapes). We predicted that interaction specifically about categories (*category-relevant discussion*) might lead to the greater coherence of L1-L2 pairs’ categories, than would discussion about unrelated information (*category-irrelevant discussion*).

In Experiment 7, we allowed all pairs to see and discuss their partner’s categories at the intervals between sorting in the interaction phase, but this time we manipulated the goal of the discussion to either reflect a non-coordinative goal (i.e., comparable to the category-relevant condition of Experiment 6) or a coordinative goal, within each L1-L2 pair. We predicted that having a coordinative goal would lead to increased L1-L2 category coherence post-discussion, compared with having a non-coordinative goal (and also that the need for explicit coordination between partners might be necessary for this type of interaction to increase category coherence between sorters in L1-L2 dialogues).

At the group level, we predicted that category-relevant interaction with one partner might lead to greater group coherence between all L1-L2 pairs, than category-irrelevant interaction (Experiment 6) – and that having a coordinative goal during interaction with one partner might increase group coherence, over having category-relevant discussion with a non-coordinative goal (Experiment 7). Again, an alternative would be that a coordinative context is necessary for category-relevant interaction with one partner to increase category coherence at all, across the L1-L2 group.

Experiment 6

2.1 Methods

2.1.1 Participants. 40 L1 monolingual speakers of British English (henceforth, L1 speakers) and 40 L1-Mandarin Chinese/L2-English bilinguals (henceforth, L2 speakers) took part in the study (Females: L1 = 31; L2 = 31). Participants were recruited through the MyCareerHub (https://mycareerhub.ed.ac.uk) and the Psychology SONA system for undergraduate students at Edinburgh (https://universityofedinburgh-ppls.sona-systems.com). Ages ranged from 18-30 years (L1: M=19.85, SD=2.20; L2: M=22.25, SD=2.93). The L2 speakers were native speakers of Mandarin Chinese (i.e., started speaking Mandarin before the age of 6 years), and began formal education in English at approximately 12 years of age. They had lived in China or Taiwan until at least the age of 16 years, and had not been resident in the UK for longer than 4 years. Participants completed an excerpt of the Language Experience and Proficiency questionnaire (see Appendix N), which focused on their learning and usage of English, in order to check participants met the criteria (Marian, Blumenfeld & Kaushanskaya, 2007). The University of Edinburgh Psychology ethics committee approved this study.

2.1.2 Stimuli. Stimuli were 128 grayscale images of common, dishware items that would generally be called ‘bowls’ and ‘plates’ in English, as well as some items that had characteristics of both (Fig. 1). We included some ambiguous items, in order to increase the likelihood of variability in participants’ linguistic categorization. We first pre-tested the items to determine their dominant names using 20 native British English speakers and 20 Mandarin-English bilinguals. From this, we selected 32 critical items for the test phases in the paired study (see Appendix O). The remaining 96 items were used in the interaction phase of the experiment (see Appendices P-S). A further 40 participants (N=20 British English monolinguals and 20 Mandarin-English bilinguals) completed a second online pre-test, in which they were asked to label each of the 32 test items using either ‘bowl’ or ‘plate’. Results were analysed using a GLMM analysis in R 3.2.1 (R Core Team, 2015), with the lme4 package, version 1.1-8 (Bates et al., 2014). The threshold for statistical significance was set at $|p| < .05$ (Baayen, 2008). These results demonstrated a significant difference between the British English monolingual and Mandarin-English bilingual naming patterns, β=-0.49, SE=0.14, p<.001 (see Fig. 2 for examples of item names across the two speaker groups). Note that all pre-testers labelled objects in English, since this was the language in which the main task would be conducted.

Fig. 1. Example stimuli showing the types of dishware objects participants categorized.

Fig. 2. Pre-test: Subset of stimuli demonstrating name agreement for more prototypical items, and disagreement for less prototypical items across L1 and L2 speakers.

Stimuli were split across eight rounds in the study. The 32 critical items were shown for the first two and the final two rounds. These 32 items were identical across the four rounds (although their positioning onscreen was randomized). The other 96 items were divided into 4 sets of 24 items and used across the interaction phase. One set was used in each of Rounds 3-6, with a different random order for each pair of participants.

We also manipulated when participants had access to labels for sorting. In rounds 1 and 8, participants sorted the items into two groups without labels. In rounds 2-7, participants were also given the labels ‘bowl’ and ‘plate’ alongside the items and asked to use these to label their categories as they sorted (Fig. 3). Labels were introduced in round 2 primarily to encourage linguistic categorization (i.e., instead of non-linguistic categorization). Therefore, our critical rounds for measuring category coherence at pre- and post-interaction were rounds 2 (pre-test) and 7 (post-test). Stimuli and labels were presented on laptops (i.e., each participant had their own laptop for item sorting). Labels remained the same for participants across rounds 2-7.

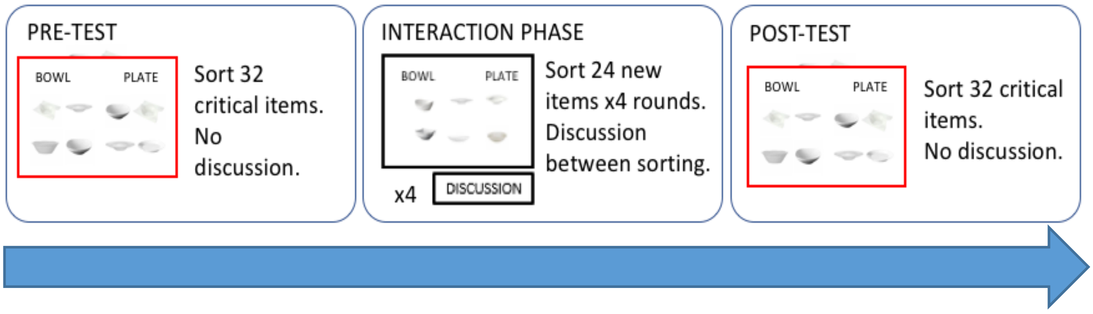


Fig. 3. Pre-test (round 2) and Post-test (round 7) sets were used to measure category coherence pre- and post-interaction between L1-L2 pairs. Rounds 3-6 formed the interaction phase.

2.1.3 Procedure. L1 speakers were randomly paired with L2 speakers to form 40 L1-L2 pairs. Pairs were randomly assigned to one of two conditions for Discussion type (i.e., category-relevant or category-irrelevant). Pairs sat opposite from one another with a barrier in place so that they could not see their partner during sorting (Fig. 4). Participants saw each set of items onscreen and were told 'Group these items into two groups in a way that makes sense to you; for example, in way that would make sense for an online store selling such items'. The categories they formed could comprise any number of items. However, they could only make two categories per round and all items had to be assigned to one of the two categories. They sorted the items by dragging items to the left or right of the screen in order to form two categories. They were given 2.5 minutes maximum to sort the items in each round.

Participants sorted items without any exposure or discussion between rounds 1, 2, 7 and 8, but had either category-relevant or category-irrelevant discussion (i.e., for up to two minutes) at the intervals between rounds 3-6. They were not allowed to alter their groups during this discussion. Participants sorted items using the labels 'bowl' and 'plate' in rounds 2-7. They did so by dragging one label to the left and the other to the right of the screen, while sorting the items under these two labels.

In the category-relevant condition, participants were allowed to see their partner's categories at intervals and were given up to two minutes to discuss their sorting strategies between each round in the interaction phase. In the category-irrelevant condition, participants did not see their partner's categories, but were given two postcards per round, which depicted images of painted landscapes, and which they had two minutes to discuss between each round in the interaction phase. In both Discussion types, the barrier was then replaced, and a new round began.

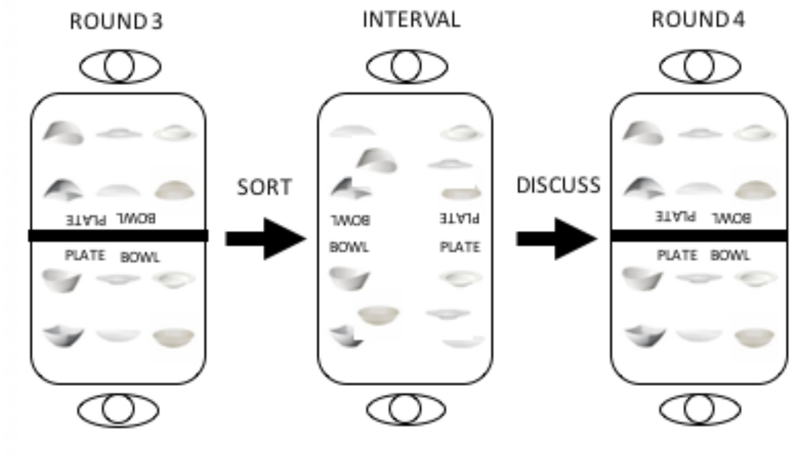


Fig. 4. Example of two rounds of sorting with discussion interval in the interaction phase (Category-relevant condition).

2.2 Results

2.2.1 Pair coherence. We analyzed category coherence for the 40 L1-L2 pairs of participants who performed the experiment together. For every participant per round, we coded whether they put each possible pair of items (32x31/2=496 item pairs) into the same category (i.e., 'bowl' or 'plate'), or not. If a participant placed two items into the same category, that item pair was coded as 1; if not, it was coded as 0. We then used this binomial data to compute a measure of association between participants who had been paired with each other in the experiment.

Descriptive statistics. In order to gain an average of pair coherence across conditions, we summed the number of item matches between a pair of participants, and divided this by the maximum number of item matches (i.e., 496). For example, if a pair of participants matched on all 496 item pairs within a round, they would be given a proportional score of 1. If they matched on no item pairs, they would be given a proportional score of 0. We averaged the resulting proportions across pairs of participants for the factors of Round (Within-pairs: Pre- and Post-test) and Discussion type (Between-pairs: category-relevant vs. category-irrelevant), as summarized in Table 1.

Table 1.

Experiment 6, pair coherence: mean pair coherence scores (SD) by Discussion type for pre- and post-test.

	Discussion	
Round	**Category-relevant**	**Category-irrelevant**
Pre-test	0.58 (.19)	0.66 (.09)
Post-test	0.46 (.21)	0.61 (.12)

GLMM analysis. We analysed the binomial results using a GLMM approach. This approach allowed us to account for random variance due to differences between participant pairs. A backwards, stepwise elimination approach was used to select factors for the final model, using likelihood ratio tests to compare models. The reference level for the analysis was Discussion type: category-irrelevant by Round: pre-test. We used category-irrelevant discussion as the reference level for Discussion type, as this reflected interaction with a lack of exposure to and discussion about categories. We used pre-test as the reference level for Round, as this reflected the pre-test state of pair coherence, before the interaction phase. The initial model included Discussion type (category-relevant vs. category-irrelevant) and Round (pre- vs. post-test) as fixed effects, as well as the interaction term between these two factors. The model also included random slopes and intercepts for Discussion type and Round by Pair. This initial model was a significantly better fit of the data than the null model ($X^2(19) = 1874.60$, $p < .001$). Removing the interaction term between factors did not produce a significantly worse fit ($p > .05$). Removing Round as a fixed effect significantly reduced model fit ($X^2(15) = 1860.20$, $p < .001$), as did removing Discussion type ($X^2(15) = 26.23$, $p < .05$). As such, the model of best fit included Discussion type and Round as fixed effects, with random slopes and intercepts for Discussion type and Round by Pair (see Tables 2 & 3). Under this model, there was a significant effect of Round, such that pair coherence was lower in the post-test than in the pre-test (β= -0.30, SE = 0.11, p <. 01). There was a significant effect of Discussion type, suggesting that having category-relevant discussion led to lower overall pair coherence than did category-irrelevant discussion (β= -0.46, SE = 0.19, p <. 05).

Table 2.

Experiment 6, pair coherence: Beta, standard errors, z and p-values for fixed effects on pair coherence. Model fit by REML.

Fixed Effects	β	S.E.	z	p
Intercept	0.72	0.09	7.92	p<.001
Discussion	-0.46	0.19	-2.37	0.02
Round	-0.30	0.11	-2.70	0.01

Table 3.

Experiment 6, pair coherence: Variance for random effects. Model fit by REML.

Random Effects		
Pair	Intercept	0.02e-01
	Discussion	0.15
	Round	0.07
	Discussion:Round	0.71

No. of Observations = 39860.

2.2.2 Group coherence. We investigated the effects of category-relevant versus category-irrelevant discussion at the group level by comparing category coherence between every possible pair of participants across the different types of Speaker pair (i.e., all possible L1-L1 pairs of participants, all L2-L2 pairs of participants and all L1-L2 pairs of participants who did not perform the task together) for Experiment 6. We compared pairs across Round (pre- vs. post-test) and Discussion type (category-relevant vs. category-irrelevant). If category-relevant interaction in L1-L2 pairs caused individuals to select more abstractable (i.e., perceptual, rather than individualistic or culturally-specific) dimensions for sorting that worked well for both the L1 and L2 speaker in a pair, then group coherence between specifically the L1-L2 pairs might be greater by the post-test, than in the pre-test, and specifically following category-relevant discussion, as opposed to following category-irrelevant discussion.

Examples of more perceptually-general strategies from the qualitative data included dimensions such as ‘height versus width’ of the item, and whether the item ‘had a rim’. This is in contrast to more culturally-specific strategies, such as the type of food item one might consume from the item (e.g., bowls reportedly used predominantly for ‘cereal’ by the L1 speakers vs. ‘rice’ for the L2 speakers), and whether the item is used for serving an individual portion of food, or for holding several servings from which a person would serve themselves (this latter point was often distinct and crucial to categorization for the L2 speakers, but not crucial for L1 speakers).

Calculating coherence across all pairs. In order to compare category coherence across every possible pair of participants within each condition, we implemented the CCM (Romney et al., 1986; Malt et al., 1999; Ameel et al., 2005; White et al., 2016). For every participant, we again coded whether they put each possible pair of items (again, 496 item pairs per round) into the same category, or not, and then used this to compute a measure of association between participants. For example, if a pairing of participants matched on all 496 item pairs within a round, they would be given a score of 1. If they matched on no item pairs, they would be given a score of 0. We calculated these proportional coherence scores for each possible pair within Experiment 6 by Round (pre- vs. post-test), Speaker pair (L1-L1, L2-L2 or L1-L2) and Discussion type (category-relevant vs. category-irrelevant). We then averaged these scores by each participant and used this proportional data for the group coherence analysis.

Descriptive Statistics. Average proportions of group coherence across Round (pre- vs. post-test) and Discussion type for all possible Speaker pairs (L1-L1, L2-L2 and L1-L2) across rounds are summarized in Table 4. Individual data points (averaged by participant) are plotted across the conditions as bee-swarm plots (see Figures 5-7).

Table 4.

Experiment 6, group coherence: mean group coherence scores (SD) by Discussion type for pre- and post-test.

Pair	Speaker status	Discussion type	Round	Average CCM	SD
L1-L1	L1	Category-irrelevant	Pre-test	0.658	0.073
L1-L1	L1	Category-relevant	Pre-test	0.642	0.142
L1-L1	L1	Category-irrelevant	Post-test	0.705	0.051
L1-L1	L1	Category-relevant	Post-test	0.547	0.158
L1-L2	L1	Category-irrelevant	Pre-test	0.682	0.075
L1-L2	L2	Category-irrelevant	Pre-test	0.682	0.066
L1-L2	L1	Category-relevant	Pre-test	0.626	0.062
L1-L2	L2	Category-relevant	Pre-test	0.626	0.134
L1-L2	L1	Category-irrelevant	Post-test	0.654	0.056
L1-L2	L2	Category-irrelevant	Post-test	0.654	0.058
L1-L2	L1	Category-relevant	Post-test	0.546	0.074
L1-L2	L2	Category-relevant	Post-test	0.546	0.131
L2-L2	L2	Category-irrelevant	Pre-test	0.727	0.064
L2-L2	L2	Category-relevant	Pre-test	0.646	0.112
L2-L2	L2	Category-irrelevant	Post-test	0.674	0.067
L2-L2	L2	Category-relevant	Post-test	0.565	0.103

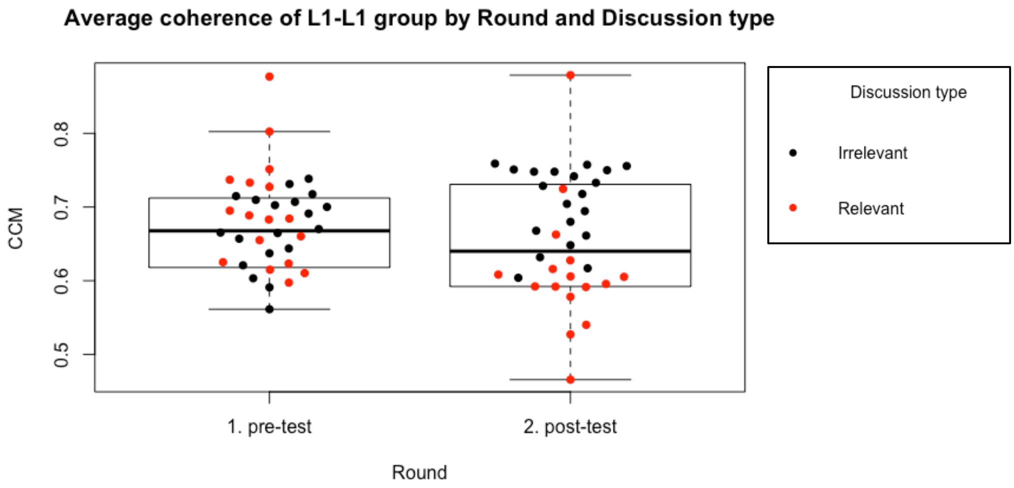


Fig. 5. Experiment 6; Average category coherence (CCM) by Discussion type (Category-relevant vs. -irrelevant) and Round (pre- to post-test) for participants in the L1-L1 Speaker pair condition (*N*=80).

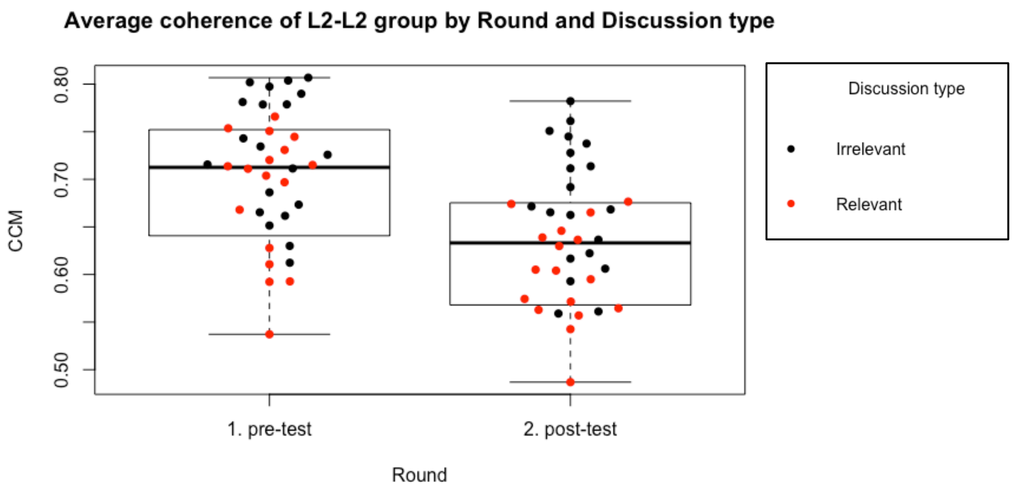


Fig. 6. Experiment 6; Average category coherence (CCM) by Discussion type (Category-relevant vs. -irrelevant) and Round (pre- to post-test) for participants in the L2-L2 Speaker pair condition (*N*=80).

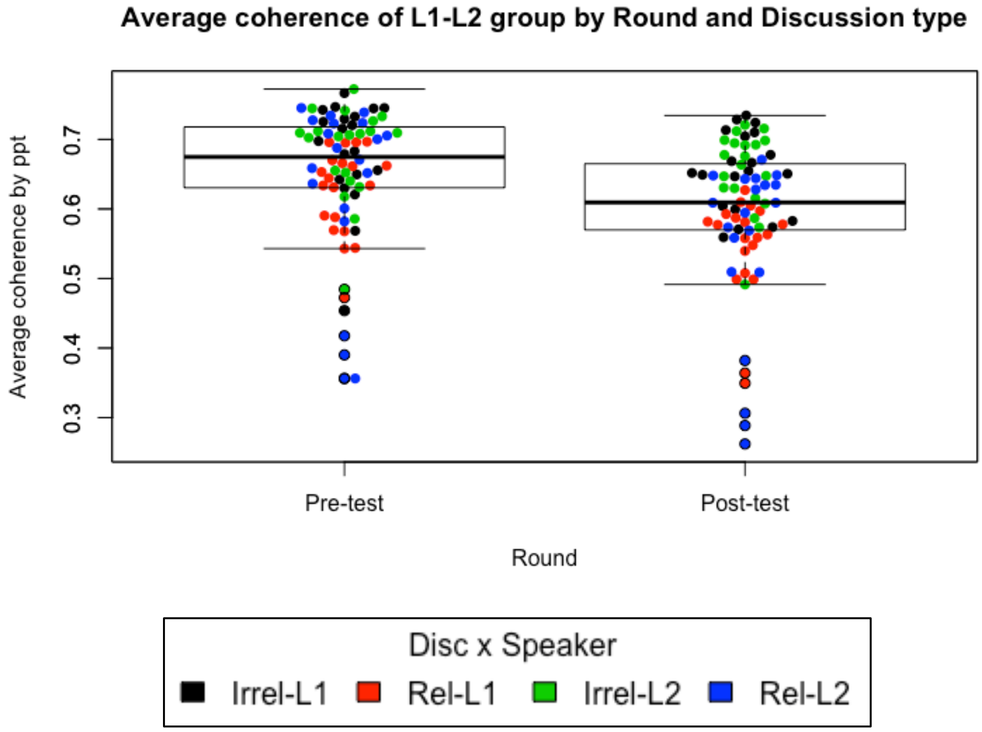


Fig. 7. Experiment 6; Average category coherence (CCM) by Discussion type (Category-relevant vs. -irrelevant) and Round (pre- to post-test) for participants in the L1-L2 Speaker pair condition (N=160).

LME analysis. In order to normalize the sampling distribution of the proportional CCM scores, they were Z-transformed with $Z=0.5*\ln[(1+r)/(1-r)]$. To test for the effects of Discussion type, Speaker pair and Round on group coherence, data were analyzed in R with a linear mixed-modeling approach again using the lme4 package (Bates et al., 2014). This approach allowed us to account for random variance due to differences between participants. The threshold for statistical significance was set at $|t| > 2$ (Baayen, 2008). A backwards, stepwise elimination approach was used to select factors for the final model, using likelihood ratio tests to compare models. Models included random slopes and intercepts for Discussion type, Speaker pair and Round by Participant. The reference level for this analysis was Discussion type: category-irrelevant by Speaker pair: L1-L1 by Round: pre-test. We again used category-irrelevant discussion as the reference level for Discussion type, as this reflected interaction with a lack of exposure to and discussion about categories, and pre-test as the reference level for Round, as this reflected the pre-test state of group coherence between pairs, before the interaction phase. We used L1-L1 pair coherence as our reference level for Speaker pair, as this represented the general level of coherence between two native speakers of English. The initial model included Speaker pair (L1-L1, L2-L2 and L1-L2), Discussion type (category-relevant vs. category-irrelevant) and Round (pre- vs. post-test) as fixed effects, as well as the interaction terms between fixed effects. This full model was a significantly better fit of the data than was the null model ($X^2(10) = 101.76, p < .001$). Removing any of the interaction terms between the fixed effects significantly reduced model fit ($p < .001$ for all). Therefore, the full model was the model of best fit (see Tables 5 & 6). Under this model, there was a significant effect of Round, such that group coherence was generally lower in the post-test, than in the pre-test (β= -0.23, SE = 0.04, t = -5.88). There was also a significant effect of Discussion type, such that pairs in the category-relevant condition had lower group coherence than pairs in the category-irrelevant discussion (β= -0.40, SE = 0.06, t = -7.00). There was a significant interaction between Discussion Type and Round, such that pairs who had category-relevant discussion had lower group coherence in the post-test, when compared to pairs who had category-irrelevant discussion (β= -0.16, SE = 0.04, t = -4.01). Importantly to addressing the effects of category relevant or irrelevant discussion on specific groups' coherence – there was a significant three-way interaction between Speaker pair, Round and Discussion type (β= 0.07, SE = 0.04, t = 2.00) (see Tables 5 & 6). To investigate this three-way interaction further, the data was subsetted by Speaker pair (L1-L1, L2-L2 & L1-L2) and analyzed with Discussion type by Round interaction-only models (with random effects for the interaction between Discussion type and Round by Participant). The significance value for these models was Bonferroni corrected to p = 0.017. Only the L1-L1 dataset demonstrated a significant interaction of Discussion type and Round, suggesting that category-relevant discussion affected the category coherence of L1 speakers as a group, but not L2 speakers as a group, or L1-L2 speakers (β= -0.33, SE = 0.13, t = -2.59) (see Tables 7 & 8).

Table 5.

Experiment 6, group analysis: Beta, standard errors and t-values for fixed effects on group coherence Z-score. Model fit by REML.

Fixed Effects	**β**	**S.E.**	**t**
Intercept	0.01	0.06	0.19
Speaker pair	-0.09	0.06	-1.35
Round	-0.23	0.04	-5.88
Discussion	-0.40	0.06	-7.00
Speaker pair:Round	-0.05	0.04	-1.28
Speaker pair:Discussion	0.02	0.06	0.32
Round:Discussion	-0.16	0.04	-4.01
Speaker pair:Round:Discussion	0.07	0.04	2.00

Table 6.

Experiment 6, group analysis: Variance and residual for random effects. Model fit by REML.

Random Effects		
Participant	Intercept	0.06e-08
	Speaker pair	0.19
	Discussion type	0.15
	Round	0.01
Residual		0.43

No. of Observations = 320.

Table 7.

Experiment 6, post-hoc group analysis (L1-L1): Beta, standard errors and t-values for fixed effects on group coherence Z-score. Model fit by REML.

Fixed Effects	**β**	**S.E.**	**t**
Intercept	0.02	0.13	0.13
Round:Discussion	-0.33	0.13	-2.59

Table 8.

Experiment 6, post-hoc group analysis (L1-L1): Variance and residual for random effects. Model fit by REML.

Random Effects		
Pair	Discussion type:Round	0.00
Residual		1.28

No. of Observations = 80.

2.3 Summary of Experiment 6

In Experiment 6, having category-relevant discussion surprisingly did not lead to increased category coherence in the post-test, compared with having category-irrelevant discussion. This suggests that category-relevant discussion alone might not be enough to support the development of greater group coherence between pairs in which interlocutors have significantly different linguistic categories to begin with. And this may be due to the different ways that their native languages map words onto referents. The group level analysis suggested a difference in effects by Discussion type across the different speaker pair groups. A subsequent post-hoc analysis confirmed that the effects of Discussion type on coherence from pre- to post-test were present in the L1-L1 group, but not the L2-L2 or L1-L2 groups. Overall, the pair and group analyses suggest that, under the conditions of Experiment 6, there is not clear evidence for a positive effect of category-relevant discussion on group coherence between L1 and L2 speakers.

Experiment 7

In Experiment 6, participants lacked a coordinative goal for sorting and results showed no significant difference in pairs' category coherence whether they were exposed to and discussed each other's categories, or not. Given the potential difficulties in the coordination of categories between speakers of different native languages, we argue that the context of the task (i.e., the goal of sorting) might be crucial to the formation of greater category coherence between L1-L2 interlocutors. As such, we conducted Experiment 7 with a manipulation of Goal type to address whether having a coordinative goal when sorting would produce greater category coherence in L1-L2 pairs following discussion, than would having a non-coordinative goal (i.e., as was the case in Experiment 6).

3.1 Methods

3.1.1 Participants. 40 new L1-monolingual speakers of British English and 40 new L1-Mandarin/L2-English bilinguals formed 40 L1-L2 pairs (Females: L1 = 30; L2 = 35). Ages ranged from 18-30 years (L1: M=21.05, SD=2.02; L2: M=23.00, SD=2.67). The criteria for the Mandarin-English participants was the same as in Experiment 6.

3.1.2 Stimuli. Stimuli were identical to those used in Experiment 6. Again, labels remained the same for participants across rounds 2-7.

3.1.3 Procedure. The procedure was identical to that of Experiment 6, except for the instructions given in relation to Goal type. In the non-coordinative condition, participants saw each other's categories and labels at intervals between rounds 3-6, and then discussed their categories with a goal of sorting in a way that made sense to the participant alone. They were told 'Group these items into two groups in a way that makes sense to you; for example, in way that would make sense for an online store selling such items' (i.e., identical instructions to Experiment 6). In the coordinative condition, participants saw each other's categories and labels at intervals between rounds 3-6, and then discussed their categories with a goal of sorting in a way that made sense to both the participant and their partner. As such, they were told 'Group these items into two groups in a way that makes sense to you and to your partner; for example, in way that would make sense for an online store selling such items'.

3.2 Results

3.2.1 Pair coherence. We investigated the effects of Goal type on the 40 L1-L2 pairs of participants who performed the experiment together. For every participant, we coded whether they put each of the 496 pairs of items into the same category or not, and used this data to compute a measure of association between participants who had been paired with each other in the experiment. This resulted in binomial data for 40 participant pairs in total.

Descriptive statistics. We averaged the resulting proportions across pairs of participants for the factors of Round (Within-pairs: Pre- and Post-test) and Goal type (Between-pairs: coordinative goal vs. non-coordinative goal), as summarized in Table 9.

Table 9.

Experiment 7, pair coherence: mean pair coherence scores (SD) by Goal type for pre- and post-test.

	Goal	
Round	**Coordinative**	**Non-coordinative**
Pre-test	0.58 (.16)	0.63 (.16)
Post-test	0.71 (.11)	0.66 (.15)

GLMM analysis. We again analysed the binomial pairwise results using a GLMM approach. This approach allowed us to account for random variance due to differences between participant pairs. The threshold for statistical significance was set at $p < .05$ (Baayen, 2008). A backwards, stepwise elimination approach was used to select factors for the final model, using likelihood ratio tests to compare models. The reference level for this analysis was Goal type: non-coordinative goal by Round: pre-test. The initial model included Goal type (coordinative goal vs. non-coordinative goal) and Round (pre- vs. post-test) as fixed effects, as well as the interaction term between these two factors. It also included random slopes and intercepts for Goal type and Round by Pair. This initial model was a significantly better fit of the data than the null model ($X^2(19) = 2308.10$, $p < .001$). Removing the interaction term between the factors did not significantly reduce model fit

($p > .05$). Of the two single factor models, a Round only model was a significantly better fit of the data than a Goal type only model ($X^2(0) = 3.25, p < .001$). The Round only model was a significantly better fit of the data than the null model ($X^2(10) = 2305.90, p < .001$). As such, the model of best fit included Round as a fixed effect, with random slopes and intercepts for Round by Pair (see Tables 10 & 11). Under this model, there was a significant effect of Round, such that pair coherence scores were greater in the post-test, than in the pre-test (β= 0.41, SE = 0.18, $p < .05$).

Table 10.

Experiment 7, pair coherence: Beta, standard errors, z and p-values for fixed effects on pair coherence. Model fit by REML.

Fixed Effects	β	S.E.	z	p
Intercept	0.46	0.12	3.84	p<.001
Round	0.41	0.18	2.29	p<.05

Table 11.

Experiment 7, pair coherence: Variance for random effects. Model fit by REML.

Random Effects		
Pair	Intercept	0.58
	Round	1.28

No. of Observations = 39860.

3.2.2 Group coherence. We investigated the effects of having coordinative versus non-coordinative goals at the group level by comparing category coherence between every possible pair of participants across the different types of Speaker pair (i.e., all possible L1-L1 pairs of participants, all L2-L2 pairs of participants and all L1-L2 pairs of participants) for Experiment 7. We compared pairs' coherence across Round (pre- vs. post-test) and Goal type (coordinative goal vs. non-coordinative goal). If category-relevant interaction in L1-L2 pairs only increases group coherence when pairs share a coordinative goal of selecting the same strategies for categorization, then group coherence between specifically the L1-L2s might be greater in the post-test for pairs who sorted with a coordinative goal, compared with pairs who sorted with a non-coordinative goal.

Calculating coherence across all pairs. Again, to make multiple comparisons for every possible pair, we implemented the CCM (Romney et al., 1986) by coding whether participants put each possible pair of items into the same category, or not, and using this data to compute a measure of association between each possible pair of participants. We calculated proportional scores for every possible pair of participants within Experiment 7 for coordinative goal versus non-coordinative goal conditions within the pre- and post-test sets.

Descriptive Statistics. Average proportions of group coherence across Round (pre- and post-test) and Goal type for all possible pairs (L1-L1, L2-L2 and L1-L2) are summarized in Table 12. Individual data points (averaged by participant within Speaker pair) are plotted across the conditions as bee-swarm plots (see Figures 8-10).

Table 12.

Experiment 7, group coherence: mean group coherence scores (SD) by Goal type for pre- and post-test.

Pair	Speaker status	Goal type	Round	Average CCM	SD
L1-L1	L1	Non-coordinative	Pre-test	0.728	0.045
L1-L1	L1	Coordinative	Pre-test	0.769	0.046
L1-L1	L1	Non-coordinative	Post-test	0.666	0.083
L1-L1	L1	Coordinative	Post-test	0.723	0.062
L1-L2	L1	Non-coordinative	Pre-test	0.671	0.111
L1-L2	L2	Non-coordinative	Pre-test	0.671	0.065
L1-L2	L1	Coordinative	Pre-test	0.651	0.038
L1-L2	L2	Coordinative	Pre-test	0.651	0.109
L1-L2	L1	Non-coordinative	Post-test	0.663	0.077
L1-L2	L2	Non-coordinative	Post-test	0.663	0.038
L1-L2	L1	Coordinative	Post-test	0.687	0.043
L1-L2	L2	Coordinative	Post-test	0.687	0.036
L2-L2	L2	Non-coordinative	Pre-test	0.693	0.114
L2-L2	L2	Coordinative	Pre-test	0.649	0.107
L2-L2	L2	Non-coordinative	Post-test	0.709	0.069
L2-L2	L2	Coordinative	Post-test	0.685	0.041

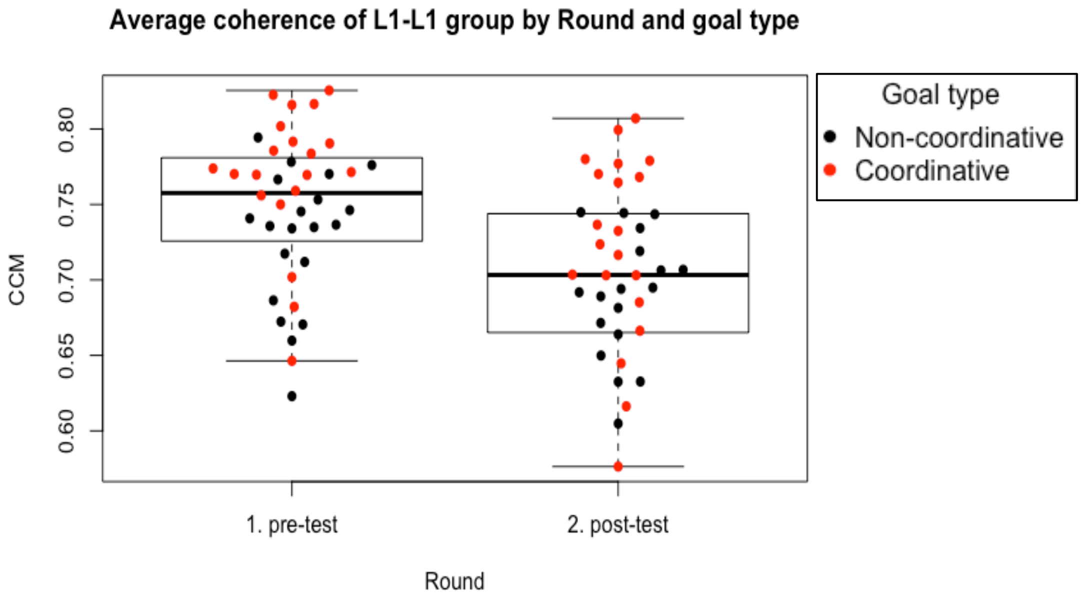


Fig. 8. Experiment 7; Average category coherence (CCM) by Goal type (Non-coordinative vs. coordinative) and Round (pre- to post-test) for participants in the L1-L1 Speaker pair condition (*N*=80).

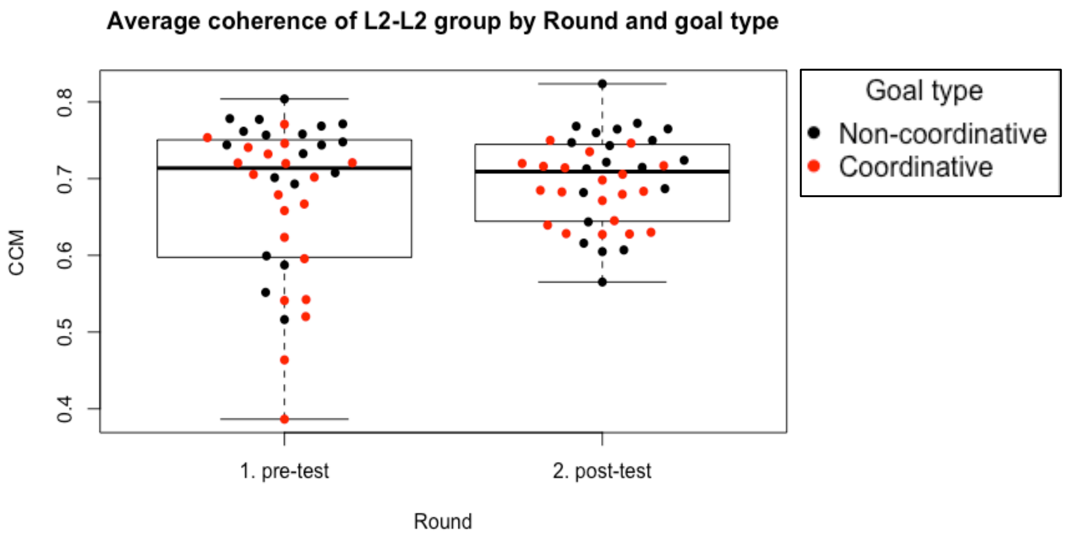


Fig. 9. Experiment 7; Average category coherence (CCM) by Goal type (Non-coordinative vs. coordinative) and Round (pre- to post-test) for participants in the L2-L2 Speaker pair condition (*N*=80).

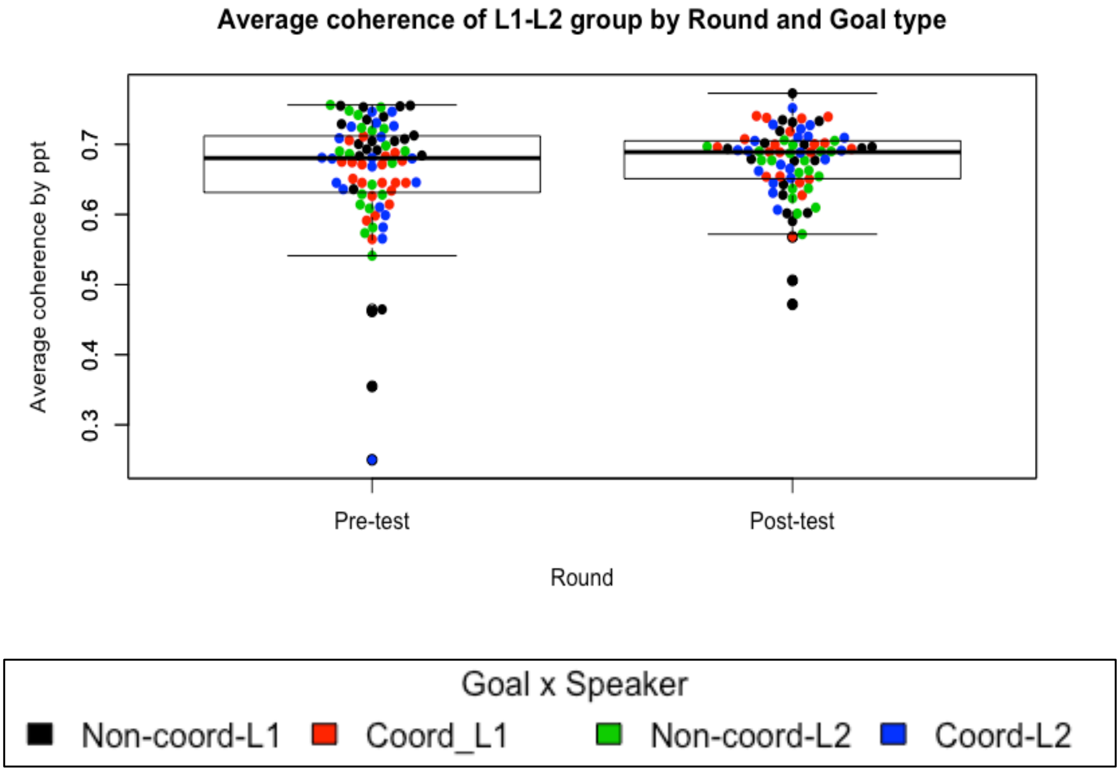


Fig. 10. Experiment 7; Average category coherence (CCM) by Goal type (Non-coordinative vs. coordinative) and Round (pre- to post-test) for participants in the L1-L2 Speaker pair condition (*N*=160).

LME analysis. Again, CCM scores were normalized with a *Z*-transformation. To test for the effects of Goal type, Speaker pair and Round on group coherence, data were analyzed using an LME analysis. A backwards, stepwise elimination approach was used to select factors for the final model, using likelihood ratio tests to compare models. Models included random slopes and intercepts for Goal type, Speaker pair and Round by Participant. The reference level for this analysis was Goal type: non-coordinative goal by Speaker pair: L1-L1 by Round: pre-test. The initial model included Speaker pair (L1-L1, L2-L2 and L1-L2), Goal type (coordinative goal vs. non-coordinative goal) and Round (pre- vs. post-test) as fixed effects. This full model was a significantly better fit than the null model ($X^2(10) = 43.47, p < .001$). Removing the interaction term between Round and the other two factors significantly reduced model fit ($X^2(3) = 14.14, p < .01$), as did removing the interaction term between Speaker pair and the other two factors ($X^2(3) = 13.47, p < .01$). However, the full, three-way model was not a significantly better fit than a model with an interaction term between Speaker pair and Round, and a separate fixed effect of Goal type ($p > .05$). Therefore, the model was the model of best fit included fixed effects of Speaker pair, Goal type and Round, with an interaction term between Speaker pair and Round only (see Tables 13 & 14). Under this model, there was a significant effect of Speaker pair, comparing L2-L2 and L1-L2 speaker pairs' coherence to that of L1-L1 speaker pairs (β= -0.27, SE = 0.05, t = -5.34). There was a significant interaction between Speaker pair and Round *(*β= 0.16, SE = 0.05, t = 3.12). However, as the model did not allow for three-way comparisons between Speaker pair, Goal type and Round, further post-hoc analyses were not conducted.

Table 13.

Experiment 7, group coherence: Beta, standard errors and t-values for fixed effects on group coherence Z-score. Model fit by REML.

Fixed Effects	**β**	**S.E.**	**t**
Intercept	0.02e-14	0.06	0.00
Speaker pair	-0.27	0.05	-5.34
Round	-0.02e-01	0.05	-0.03
Goal	0.03e-01	0.06	0.52
Speaker pair:Round	0.02	0.05	3.12

Table 14.

Experiment 7, group coherence: Variance and residual for random effects. Model fit by REML.

Random Effects		
Pair	Intercept	0.10
	Speaker pair	0.00
	Goal type	0.01
	Round	0.02
Residual		0.79

No. of Observations = 320.

3.3 Summary of Experiment 7

In Experiment 7, we introduced a shared goal of coordination between partners to test whether the need for coordination between partners is necessary for increased category coherence between interlocutors in L1-L2 dialogues. However, category-relevant discussion with a coordinative goal did not result in greater pair or group coherence for L1-L2 pairs (or L2-L2 pairs) than did category-relevant discussion with a non-coordinative goal.

General Discussion

Across two experiments, 80 L1-L2 pairs individually categorized dishware across eight rounds, and interacted with their partner at intervals between rounds 3-6. Firstly, we focused on L1-L2 pairs' category coherence pre- and post-discussion. Discussion of categories alone did not increase the coherence of pairs' categories (*pair coherence*) in Experiment 6, and neither did coordinative discussion in Experiment 7. Additionally, we examined coherence across all participants within each condition (*group coherence*). Discussion of categories differentially affected the similarity of L1 speakers' categories to those of other L1 speakers, but did not affect the similarity of L2 speakers' categories with other L2 speakers, or the similarity of L1 and L2 speakers' categories to each other (Experiment 6). The addition of a coordinative goal with discussion of categories did not increase group coherence, over discussion of categories with a non-coordinative goal (Experiment 7).

4.1 Pair coherence

Previous research has posited the importance of exposure to and discussion about a person's categories in learning how that person conceptualizes and categorizes items, and hence in developing similar categories to theirs (Markman & Makin, 1998). In addition, research on reference in dialogue suggests that interlocutors tend to coordinate on the labels that they would give to certain objects during interaction, hence also coordinating their linguistic categories for those objects (e.g., Brennan & Clark, 1996; Garrod & Doherty, 1994). However, this research focused on discussion between native speakers of the same language who already had significant overlap in the way that they labelled and, therefore, linguistically categorized objects. In the current study, we examined whether interaction led to greater category coherence in cases where interlocutors had significantly different linguistic categories for objects. We predicted that the development of category coherence through discussion might be attenuated between L1 and L2 speakers, and that a goal of coordination might be necessary in order for interaction to increase category coherence in L1-L2 dialogues.

The pairwise results of Experiment 6 supported the prediction that interaction about categories might not lead to greater category coherence between speakers with significantly different linguistic categories, showing that category-relevant discussion did not increase pair coherence for L1-L2 pairs from pre- to post-interaction. In fact, participants in the condition of category-relevant discussion overall had lower pair coherence, than did participant in the category-irrelevant discussion condition. Some research has found that interaction and, so, negotiation about categories sometimes led to lower category coherence and less optimal category structures between pairs of sorters, than did the absence of interaction and exposure between sorters (Silvey, 2014; Thompson et al., 2014). As such, when two individuals start with significantly different linguistic categories, it may be that discussion of and exposure to a partner's categories alone is not sufficient to increase category coherence, and – in some contexts – it can lead to the divergence of interlocutors' categories.

In Experiment 7, we manipulated the goal of a pair's discussion, such that half of the pairs were given a coordinative goal of attempting to sort the items in a way that would make sense them and also to their partner. The other half of participant pairs were given the goal of sorting in a way that made sense to them alone (i.e., as was the case in the category-relevant discussion of Experiment 6). If L1-L2 pairs were able to coordinate their categories and, so, increase their category coherence following discussion with a coordinative goal, this would suggest that the goal of a discussion is crucial to the development of greater pair coherence in situations where interlocutors do not have the same linguistic categories. However, while category coherence increased from pre- to post-test overall, specifically having a coordinative goal with category-relevant discussion did not yield a benefit for category coherence, over having category-relevant discussion with an individual goal, in Experiment 7. Therefore, our results suggest that having a shared goal of using the same category dimensions is alone not sufficient to overcome the initial linguistic category differences between L1 and L2 speakers (Malt et al., 1999; Ameel et al., 2005; Costa et al., 2008).

4.2 Group coherence

Secondarily, we calculated group coherence for all L1-L1, all L2-L2 and all L1-L2 pairs within each experimental condition (i.e., by discussion type in Experiment 6; by goal type in Experiment 7) to reflect the groups of: L1-L1 speaker pairs, L2-L2 speaker pairs and lastly L1-L2 speaker pairs. We did so to examine whether discussion of categories with a specific partner would also affect category coherence across people as a wider group. Specifically, we asked whether interaction with one L1 or L2 partner would yield benefits in group coherence for L1-L2 pairs that did not actually interact with each other during the task. However, the results of this group level analysis have different implications to those from the pairwise analysis. The effects of interaction at the group level relate to the selection of dimensions for categorization that are deemed acceptable not only across two individuals, but also which might be those most likely to be deemed acceptable across a group of people. This is especially so when those dimensions are based on features shared across people, such as perceptual features (Rosch & Mervis, 1975; Johnson, 1987). And by being exposed to a partner, it is possible that we are more likely to stick to such dimensions across the task and, thus, develop categories that are more similar to other people (i.e., people who interacted with another person, but not with us per se). Relating this to the development of conventions, it is possible that – since linguistic conventions (such as reference usage, Brennan & Clark, 1996) are acquired and maintained mainly through one-on-one interactions (i.e., conversations) – the same mechanisms governing conventions in one-on-one interactions might pertain to some extent to the broader, social context (i.e., how groups develop language conventions, Garrod & Doherty, 1994).

For the group level analysis in Experiment 6, we found differential effects of discussion type on category coherence from pre- to post-test, dependent on the speaker group in question. Post-hoc analyses confirmed that this change in coherence across the manipulation of discussion type occurred for the L1-L1 speaker group specifically, but not for the L2-L2 or L1-L2 speaker groups. These results might suggest an effect on sorters' similarity to others that is affected by a person's language status and, in this, that the L1 speakers may have shifted their categories more than the L2 speakers (e.g., perhaps due to accommodation; Arthur et al., 1980)

Unlike the group level results for Experiment 6, L1-L2 and L2-L2 pairs in Experiment 7 did not show differential change in group coherence due to the type of discussion had, compared with the L1-L1 pairs. Across both experiments, then, results are inconsistent on the effects of category-relevant discussion and the goal of the discussion on category coherence between L1 and L2 speakers. However, the differences in effects in Experiment 6 do suggest that the effects of discussion on coherence might be differentially affected by speaker status, and that more research is required in order to better understand how to reinforce the development of greater category coherence between speakers with significantly different linguistic categories.

4.3 Conclusion

Speaking different native languages can lead to people having different systems for categorization, and bilinguals demonstrate reliably different sorting patterns in their L2 language, compared with L1 speakers' patterns of sorting. In these Experiments, our findings demonstrate the challenges faced in developing greater category coherence between speakers with significantly different linguistic categories, even when the goal of the interaction is to achieve more similar categories to a partner. This is contrast to the majority of research on the development of coherence between native speakers of the same language, in which greater coherence seems to develop on a relatively automatic and implicit basis. Results also suggest that the effects of category-relevant discussion on category structure and coherence are affected by the status of the speaker, on the basis of whether they are an L1 or L2 speaker of the language. Lastly, they show that explicit coordination does not always lead to increased category coherence between pairs in L1-L2 dialogues.

Chapter 6

Discussion and summary

In this thesis, I aimed to examine the effects of both labelling and interaction on the similarity of people's categories, as a way to examine their shared understanding of the world (i.e., by the way in which they group objects in the world together). Information from and exposure to other people's categories can help shape (and, so, change) our understanding of how the world is divided up into categories, and make us more similar to each other. In this way, labels (i.e., conventionalized word labels) not only help us form and share these categories, but can also alter the way in which we categorize, and make categories more coherent across groups of people.

Following the results of Experiment 1, I chose to focus on the effects of labelling specifically, comparing the similarity of people's categories for novel, morphed triangle stimuli when they sorted with and without novel, non-word labels (Chapters 3 & 4). I also began examining the possibility of more general effects of labelling and interaction on the coherence of people's categories beyond the pairs of participants who performed the task together (i.e., group coherence). This allowed me to examine whether these factors changed people's categorization in a way that shifted each participant's categories closer to that of an average way of sorting (Chapters 3-5).

In this chapter, I will summarize the overall findings of this thesis, as well as the limitations and possible extensions of the methodologies and analyses used. Discussion for Experiments 1-5 will focus mainly on the potential effects of novel labels and interaction on interlocutors' categories. Discussion of Experiments 1-3 will help me draw some conclusions on the effects of exposure and interaction between people on the coherence of their categories, with and without novel labels. Discussion of Experiments 4 and 5 will add to this, in allowing me to make inferences about how having a coordinative context might affect label use in relation to category coherence across sorters. Lastly, discussion of results for Experiments 6 and 7 will switch to focusing mainly on the effects of interaction on linguistic category coherence in L1-L2 dialogues, in order to draw some conclusions regarding whether interaction can increase category coherence in cases where speakers have significantly different patterns of categorization, as a result of the different native languages that they speak.

1.1 Experiment 1

Past research has suggested that both interaction about categories and using conventionalized labels (i.e., words) when categorizing stimuli can increase people's category coherence (e.g., Markman & Makin, 1998). However, other evidence has suggested that labelling effects might not be limited to conventionalized labels (Lupyan et al., 2007; Lupyan & Casasanto, 2014). As such, in Experiment 1 I investigated whether novel labels increased category coherence between people, and how this compared to the effects of interaction (i.e., dialogue between partners) on category coherence. I asked pairs of participants to sort six sets of novel, morphed triangle stimuli into groups with or without novel, non-word labels – and with varying degrees of exposure and interaction between partners. Firstly, pairs in the Non-exposed condition sorted the shapes without labels and without seeing their partner's categories. Secondly, for pairs in the three Exposed conditions I manipulated the types of information available to the pairs: (1) pairs sorted with exposure to a partner's categories, but without labels or dialogue, (2) pairs sorted with exposure to a partner's categories and used novel, non-word labels to sort, and (3) pairs sorted with exposure and dialogue, but without the use of labels.

Surprisingly, results demonstrated that pairs of participants who sorted with novel labels (Exposed-with-labels) had greater category coherence than pairs who sorted without labels but with full dialogue (Exposed-with-dialogue). There were no significant differences in pairs' category coherence between any of the other conditions. In addition, how participants in a pair applied each of the labels to specific items did not affect how similar a pair's categories became. I argued that this was evidence for a direct effect of labelling on categorization, in which the act of labelling itself (rather than the content of the label) influenced people to select more similar category dimensions to each other (Lupyan, 2008). There was also no evidence of an increase or decrease of category coherence over time, again suggesting that the labels produced a state of greater coherence, rather than through a process by which labels gradually increased coherence across rounds.

1.2 Limitations and future directions. While the significant difference between the conditions of labelling versus dialogue led me to focus on labelling in subsequent experiments, the lack of significant differences in coherence between any of the other conditions in Experiment 1 was surprising. For example, while the results suggested that labels produced significantly greater pair coherence than did dialogue between partners, the lack of difference between the labelling condition and the non-exposed condition suggested that the categories of pairs sorting with exposure and labels were no more coherent than what was essentially a baseline condition (i.e., since participants were never exposed to one another and sorted without labels in the non-exposed condition). It could be that these different conditions do not produce significant differences in coherence, but there may also be an inability in the analysis to detect differences across the other conditions, due to a lack of power across the numerous conditions that were ran as part of the study. This is an issue I addressed in later experiments, through both the design of the experiment (e.g., number of trials; type of data collected) and the number of participant pairs per condition.

Secondly, the significant difference between the labelling condition and the dialogue condition was difficult to interpret, since neither of these conditions could be considered a baseline to one another (i.e., exposure with labels is not a baseline in comparison to exposure with dialogue, and vice versa). As such, within the design of Experiment 1 it was impossible to say whether dialogue hindered the development of category coherence in comparison to labelling without dialogue, or whether the act of labelling supported the development of greater category coherence, beyond that of dialogue. In order to deal with such issues of interpretation, conditions should be based around a baseline condition which can be used as a point of reference for comparison with all other conditions (as was the case in Experiments 2-5).

To better understand the effects of novel labels on category coherence across people, Experiments 2-5 focused on category coherence specifically under conditions with and without novel labels, whilst limiting interaction between sorters. However, another aspect to explore in the future could be to examine the effects of novel labels on category coherence within dialogue settings (e.g., similar to the paradigm of Markman & Makin, 1999), by asking pairs of participants to establish shared reference to objects using novel labels – instead of conventionalized word labels – during a joint task.

2.1 Experiments 2 and 3

Because of the effects of labelling in Experiment 1 (i.e., in contrast to dialogue), I next chose to focus on the effects of labelling, comparing the similarity of people's categories for novel, morphed triangle stimuli when they sorted with and without novel, non-word labels. Here, I also began examining the possibility of more general effects of labelling and interaction on the coherence of people's categories beyond the pairs of participants who performed the task together (i.e., examining group coherence, as well as pair coherence). This allowed me to examine whether the factors of labelling and exposure changed people's categorization in a way that shifted each participant's categories closer to that of an average way of sorting.

In Experiments 2 and 3, results did not show an effect of labels on category coherence and there was also no interaction of labels with exposure, suggesting that labels did not directly affect categorization within this paradigm. Instead, results supported an account in which exposure to others' categories is sometimes necessary for category coherence, because it suggests that people generally needed to see each other's categories in order to sort items more similarly. As such, having exposure to other people's ways of sorting might relate to the selection of more generalizable and abstractable dimensions, increasing sorters' tendencies to utilize perceptual features for categorization, over more individualistic choices for sorting the items.

2.2 Limitations and future directions. While we did not find an effect of labels for both label types, a potential issue here is that every pair of participants used the same pair of either non-word or coloured tag labels across Experiments 2 and 3, similarly to the designs of Lupyan et al. (2007) and Lupyan and Casasanto (2014). I intended to control for non-arbitrariness by avoiding particularly iconic sound symbolic differences across the non-words, and by avoiding picking combinations of colours with clear or opposing associations in Western culture (e.g., not using green and red and these might be used to represent 'right' and 'wrong' respectively). However, other labels may have produced different results on category coherence. As such, in order to investigate general effects of labelling on categorization, it would have perhaps been better to vary the novel labels across each pair of participants. This is a step I took in Experiments 4 and 5 (i.e., varying the novel labels across every participant in the with-labels conditions).

Lastly, a combined analysis of the data from Experiments 2 and 3 was not conducted due to temporal confounding of when the experiments were ran, but in the future it would be useful to attempt to directly compare the effects of linguistic and non-linguistic labels on coherence within the same experimental setting.

3.1 Experiments 4 and 5

In Chapter 4, I asked whether novel labels consistently affected people's categories only when these labels were used in categorization within coordinative contexts (i.e., in contexts in which a label was used to convey information about the categories to another person). Previous research had suggested that people develop more similar categories when they use novel labels to categorize, than when they do not use labels, in categorization tasks. However, because these tasks were paired (e.g., Experiment 2), or involved some aspect of coordination with a pre-defined category structure (Lupyan et al., 2007; Lupyan & Casasanto, 2014), we did not know whether the category coherence yielded by novel labels was general to categorization, or specifically an effect of communication.

To investigate this, I utilized an online sorting paradigm in which participants sorted greyscale images of mountains into two groups with or without novel, non-word labels, and with or without a coordinative context. I chose images of mountains because participants were unlikely to have strong preconceptions about how groups of mountains should be categorized. I also varied each pair of novel, non-word labels used by participants in the labelling condition in order to reduce the aforementioned issues of non-arbitrariness for non-words.

Experiment 4 did not demonstrate an effect of labels on category coherence across sorters. Experiment 5 showed greater category coherence for participants who sorted using novel labels, than those who sorted without labels. These results provided evidence that novel labels can affect categorization in a way that increases the potential for people's categories to overlap. However, the null results of the combined data (contrasting Experiments 4 & 5 directly) suggested that more research is needed to fully understand the relationship between labelling and the context of sorting.

3.2 Limitations and future directions. Experiments 4 and 5 both used three sets of mountain images as stimuli, across three rounds of sorting. In order to be able to be able to compare all participants across these rounds, I kept the order of set presentation the same across all participants (i.e., I did not counterbalance set order). Additionally, in previous experiments I had included round as a predictor, but failed to find any strong evidence of trends in relation to labelling and interaction effects on category coherence across time – this was another reason why I chose to not include round as a predictor in Experiments 6 and 7. However, in the future it might still be useful to try and examine the effects of round within the paradigm used in Experiments 6 and 7 (i.e., to examine whether there are any trends in coherence across time, in relation to labelling and the context of sorting).

4.1 Experiments 6 and 7

Lastly, I investigated what happened to people's categories when speakers with significantly different ways of labelling objects interacted with each other, and whether in such cases, these speakers could learn each other's ways of labelling and categorizing everyday objects through interaction. As such, in addition to examining how interaction affected category coherence between speakers of the same language, in Chapter 5, I examined the effects of interaction and goal on category coherence between monolingual and bilingual speakers of English with significantly different linguistic categories. This also meant that my focus in Experiments 6 and 7 was on the effects of interaction, more so than the effects of labelling (the latter which was the focus of my research in Experiments 2-5).

Specifically, I investigated (a) whether discussion increases the similarity of people's categories (*category coherence*) in L1-L2 pairs, (b) how the need for coordination between partners affects this process, and (c) whether these effects lead to differences in category change across L1 versus L2 speakers. L1-L2 pairs individually categorized dishware with intermittent interaction: in Experiment 6, participants discussed their categories, or unrelated images; in Experiment 7, all participants discussed their categories, but some pairs did so with a coordinative goal. Discussion of categories alone did not increase the coherence of pairs' categories (pair coherence), regardless of whether participants sorted with a coordinative goal in mind. Additionally, I examined coherence across all participants within each condition (group coherence). Discussion of categories differentially affected coherence for the similarity of L1 speakers to other L1s, but did not affect the L2-L2 or L1-L2 groups' category coherence (Experiment 6). Having a coordinative goal during category-relevant discussion did not increase group coherence, more so than category-relevant discussion without a coordinative goal (Experiment 7). I, therefore, argued that the effects of category-relevant discussion on category structure and coherence are affected by the status of the speaker and, secondly, that explicit coordination does not always lead to increased category coherence between pairs in L1-L2 dialogues.

4.1 Limitations and future directions. Although the focus of Experiments 6 and 7 was to address interaction between L1 and L2 speakers specifically, the basis for alignment (as a mechanism possible of supporting the development of category coherence) was informed by literature that focused on L1-L1 interactions (e.g., Clark & Brennan, 1991; Markman & Makin, 1998; Pickering & Garrod, 2004). While I was able to examine group coherence across all possible L1-L1, L2-L2 and L1-L2 pairs, all of the actual pairings within the task setting comprised of one L1 and one L2 speaker. As such, it would be valuable to run these experiments with additional speaker pairings, such that I had pairwise coherence results for L1-L1 dialogues and L2-L2 dialogues, within the same task setting used for the L1-L2 dialogues. In this way, the L1-L1 pairs might then be used as a direct reference level for pair coherence, against the L1-L2 pairs.

Another approach might be to employ a confederate participant to take part with each naïve participant. This would mean that I was no longer investigating legitimate interactions between naïve L1 and L2 speakers (albeit in a lab setting). However, it would allow greater insight into how individual factors might affect specific speakers' abilities to coordinate categories with a scripted L1 or L2 confederate. In this way, individual factors – such as an L2 speaker's proficiency, or language exposure – could be used as a factor to predict the individual's ability to achieve category coherence with a specific L1 partner who would behave in a predictable manner.

Similarly to Experiments 2 and 3, a combined analysis of the data from Experiments 6 and 7 was not conducted due to temporal confounding of when the experiment were ran. Again, in the future it would be useful to attempt to directly compare the effects of discussion of categories with and without a joint goal on coherence within the same experimental setting.

5.1 Conclusions and additional comments

Throughout this thesis I have argued that successful communication relies on coordination between people, and that both interaction with others about our categories and having access to labels for categories can – in certain contexts - help us to achieve greater coordination. I chose to focus on category coherence as a measure of coordination, because understanding how we refer to things in the world is a crucial component of having a shared understanding of that world. I argued for and added evidence to the proposal that labelling not only allows us to label items in the world, but that the act of labelling itself can influence how we process and group items together, in both infancy (Waxman & Markow, 1995; Althaus & Mareschal, 2014; Althaus & Plunkett, 2016), and in adulthood (Lupyan, 2008; Lupyan et al., 2007; Lupyan & Casasanto, 2014). However, my results also showed that labelling effects on coherence are not consistent across all categorization settings.

Experiment 1 provided initial evidence of the labelling advantage for category coherence in cases where participants instantiated their own novel categories − but this was only in comparison to a condition in which pairs of participants engaged in full dialogue between sorting. As such, in Experiments 2 and 3, I chose to focus on the effects of novel labels, whilst limiting interaction between sorters (i.e., exposure to each other's categories, but no dialogue). Surprisingly, labels were not beneficial to greater category coherence within these experiments. Instead, exposure to another person's way of sorting appeared crucial to the development of greater category coherence with other people. However, under such a paradigm, each participant was still influenced to some extent – and especially in the exposed conditions – by the participant with which they were partnered.

As such, in Experiments 4 and 5, I studied labelling effects on category coherence in people who performed the task as individuals, not pairs. In Experiment 4, there was no labelling advantage for coherence. This led me to reason that, given the communicative nature of words, the context in which one uses labels might be crucial to the labelling advantage for category coherence. As such, in Experiment 5, I implemented the same sorting paradigm, but changed the instructions to reflect a coordinative context in which each participant sorted the items into groups with another person's categories in mind. In Experiment 5, there was a labelling advantage for coherence, compared with sorting without labels. In summary, Experiments 1-3 did not give a clear indication of labelling effects on category coherence – perhaps due to the paired nature of the task - but Experiments 4 and 5 specifically helped to reinforce the theory that labels can directly affect categorization in some contexts, by influencing which dimensions sorters select for categorization (Lupyan, 2008). And in doing this, labels make people select dimensions more similarly to each other, increasing category coherence and, therefore, their shared understanding of how objects in the world should be parsed up. However, further work is required in order to directly compare the effects of context (coordinative vs. non-coordinative) with those of labelling on category coherence across people.

Of course, this does not mean that interaction and exposure between people are not important to the way that we learn and coordinate our understanding of the world, given that this is how we learn to label and categorize objects from infancy. While the pairwise results for exposure across Experiments 2 and 3 were inconsistent, exposure increased category coherence for groups of sorters in both of these experiments. Additionally, the results of Experiment 5 suggested that the labelling advantage for coherence was dependent on sorters having a communicative context in which they must coordinate their categories. These results, therefore, support the notion of language (and labels) as primarily communicative in purpose.

In Experiments 6 and 7, I chose to examine specifically the effects of interaction on category coherence for items with existing labels (i.e., 'bowl' versus 'plate'), within a setting where participants had significantly different linguistic categories. However, the results of Experiments 6 and 7 suggested that – in some situations – the discussion of categories alone is not be enough to increase category coherence between interlocutors. Even sorting and interacting with a coordinative goal in mind did not lead to consistent effects of increased category coherence at the pair and group levels. This is contrast to the majority of research on the development of coherence between native speakers of the same language, in which greater coherence seems to develop on a relatively automatic and implicit basis. Results also suggested that the effects of category-relevant discussion on category structure and coherence are affected by the status of the speaker, on the basis of whether they are an L1 or L2 speaker of the language.

In conclusion, all of these effects appear dependent on context. Language is strongly tied to communication and, as such, language is a tool that helps people to coordinate. In contexts in which we do not need to coordinate, novel labels do not appear to yield benefits for category coherence. Similarly, in dialogues where interlocutors do not share the same linguistic categories for items, additional effort may be needed to reinforce the development of greater category coherence between interlocutors, potentially through more explicit and less automatic routes than we might expect for interlocutors with the same linguistic categories (Costa et al., 2008).

Throughout this thesis, I suggested that factors such as interaction and labelling might affect categorization and increase coherence by influencing sorters to pick more abstractable, perceptually-shared dimensions for sorting. However, by abstractable I do not mean I assume an abstraction (or prototype) based account of categorization only. On the contrary, I believe that factors like interaction and labelling could also influence coherence by pushing sorters to select an exemplar with features that are more easily abstractable across a range of items (i.e., in contrast to them forming a prototype of more abstractable features). Therefore, I agree that either account could explain categorization and, so, interact with the effects of interaction and labelling on categorization.

I examined the effects of labelling in coordinative and non-coordinative contexts for categories formed of both naturalistic (i.e., mountains − Experiments 4 & 5) and non-naturalistic (i.e., morphed triangles – Experiments 1-3) items. Both of these stimuli types showed some labelling benefits for category coherence. Additionally, I examined the effects of interaction with naturalistic (i.e., dishware – Experiments 6 and 7) and non-naturalistic (i.e., morphed triangles – Experiments 1) items, and found differences in the results for coherence. As such, in the future it would be valuable to also investigate the effects of explicitly coordinative interaction on the coherence of people's categories for non-naturalistic items.

In terms of the methods used in this thesis, I chose to limit categorization to contexts which were highly constrained in regards to interaction between sorters and the labels that they could use for their categories. I chose such constraints to try and control for one factor, while testing for the effects of the other (e.g., to examine the effects of labelling while minimizing the effects of interaction with/exposure to others, and vice-versa). While this approach yielded some interesting results on the effects of individual factors, a more ecologically valid approach is, of course, to use environments in which sorters employ both language and interaction in the same setting in order to learn how to categorize new items (as well as to learn what to call those items). As such, I admit that in the future it would be interesting to build upon my thesis work to examine category development and coherence in contexts where sorters can utilize both dialogue within the interaction and novel labels to form their categories.

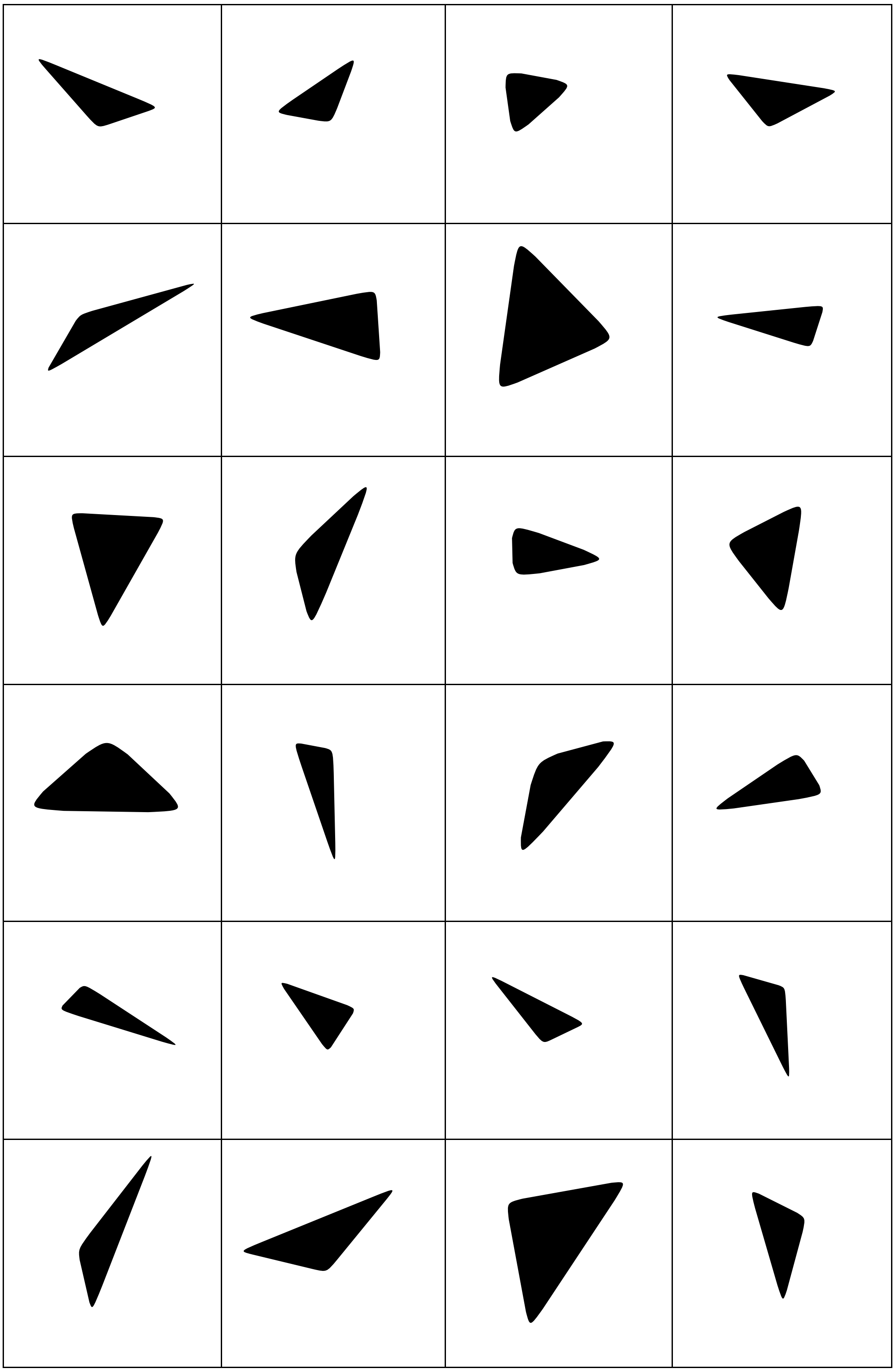
Similarly, the approach I took to categorization might be described as clustering in contrast to other methods whereby sorters see the items they have to categorize on a one-by-one basis (i.e., individuation), and it is true that people do not generally experience categories as a full item set in the real world (i.e., they more often encounter items on a one-by-one basis). However, I chose to use this clustering approach - or to let my participants see the entire set of stimuli per round - as I believed it would allow participants to better select dimensions that would work well across the range of items. I also believe that individuation is better suited to tasks which test the learning of pre-specified categories, whereas I wanted to give my participants the freedom to sort how they wanted, and then to see how different factors influenced their choices of dimensions across the full item set.

Lastly, I chose to use subordinate categories rather than basic categories as I felt that utilizing basic categories with adults would produce potentially less variation across sorters (i.e., due to the well-established nature of basic categories across speakers), from which my manipulations of labels and interaction would produce smaller, less detectable effects. But again, it would be interesting to repeat the tasks used in this thesis for basic categories to compare the amount of change and coherence in people's categories on the basis of interaction, labels and the sorter's language status (e.g., L1 vs. L2 speaker).

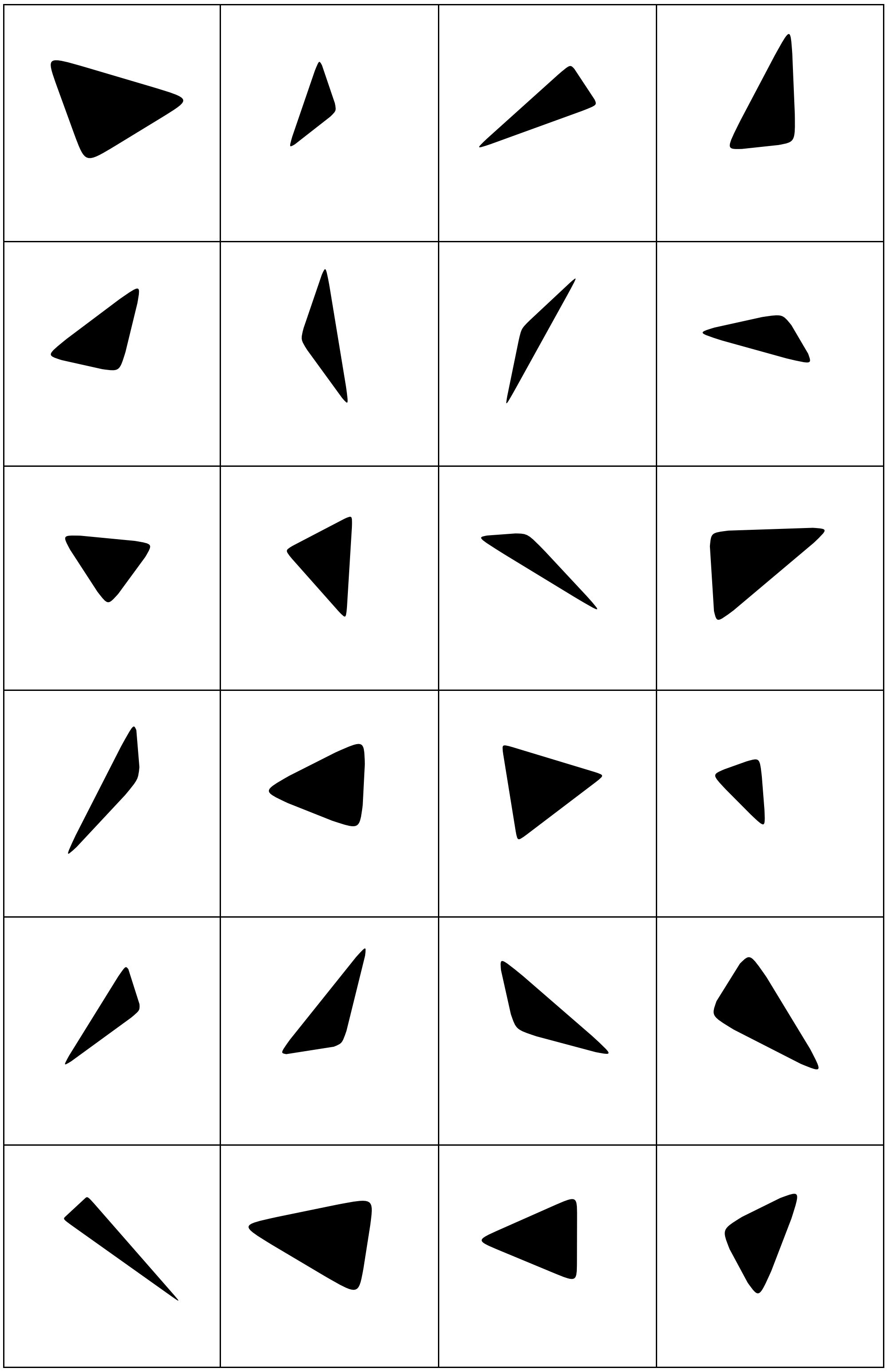
In conclusion, achieving coherence in our representations of categories can be crucial to successful communication. This is the case for both speakers of the same native language, and for speakers of different native languages with very different linguistic categories. The results of this thesis support taking a multifaceted approach to how language and interaction bolster our coordination. In addition, the results suggest that the context of coordination itself is crucial to the effects of both labelling and interaction on the coherence of our categories.

Appendices

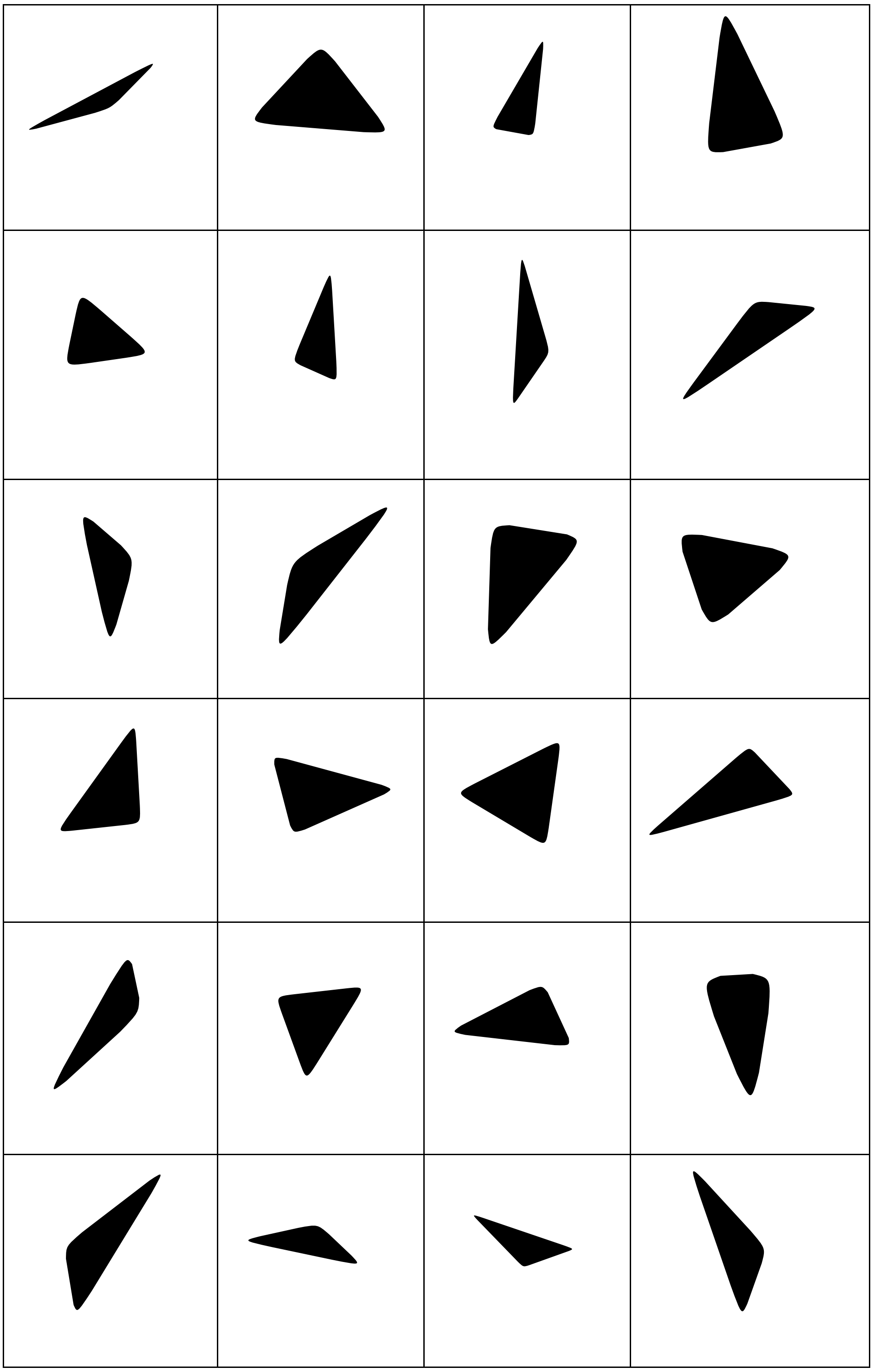
Appendix A – Experiment 1: Stimuli set



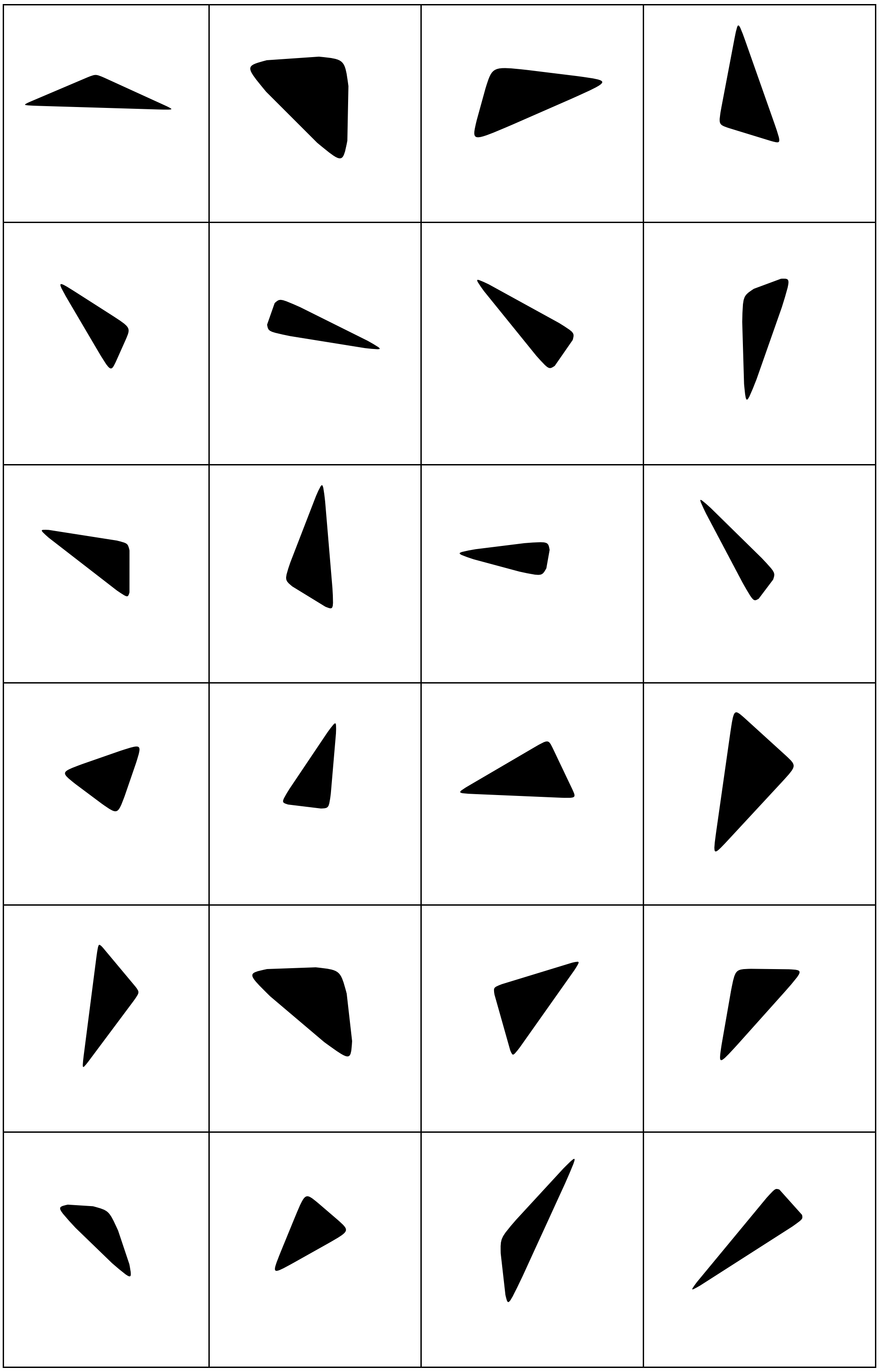
Appendix B – Experiment 1: Stimuli set B



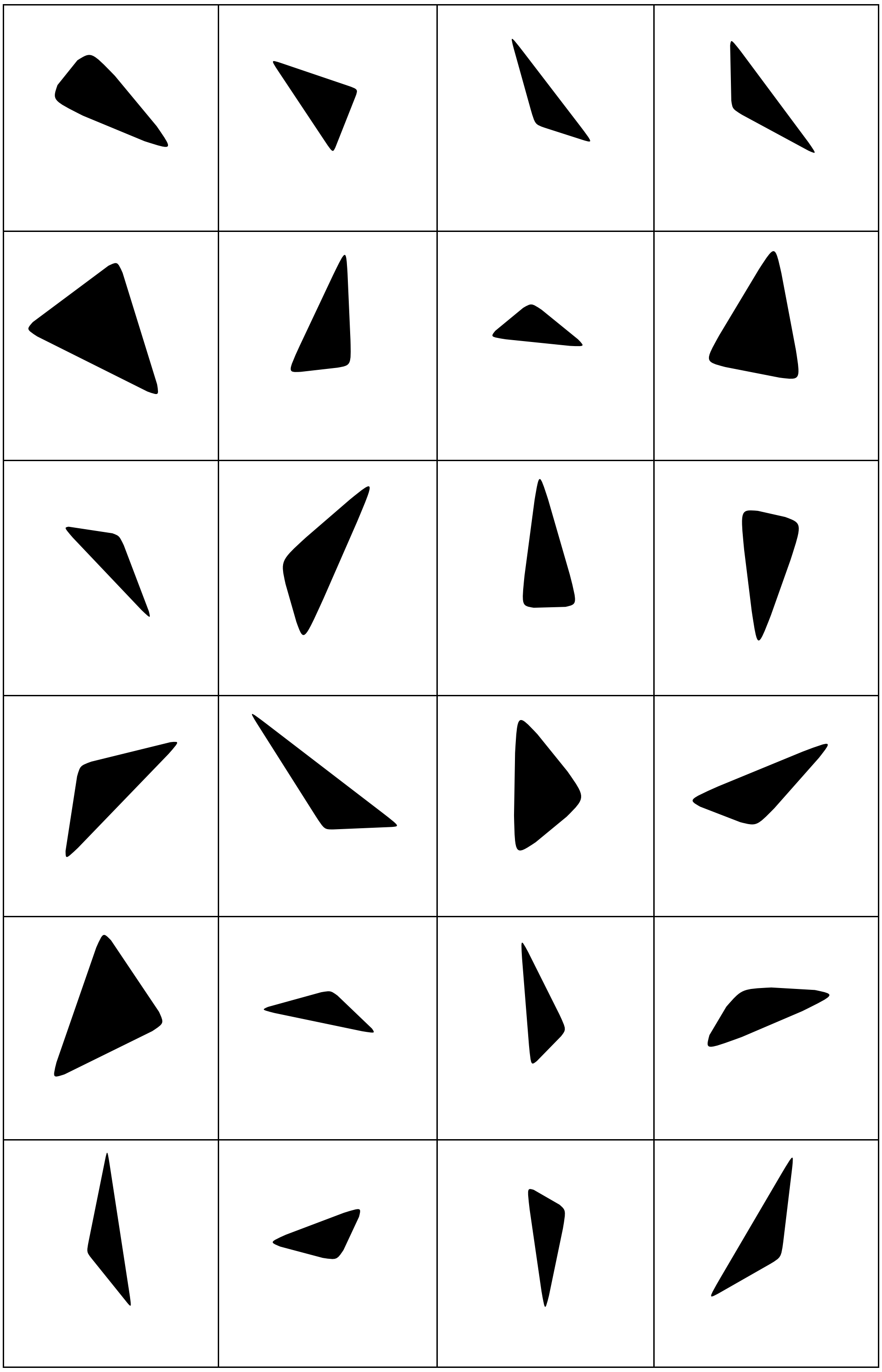
Appendix C – Experiment 1: Stimuli set C



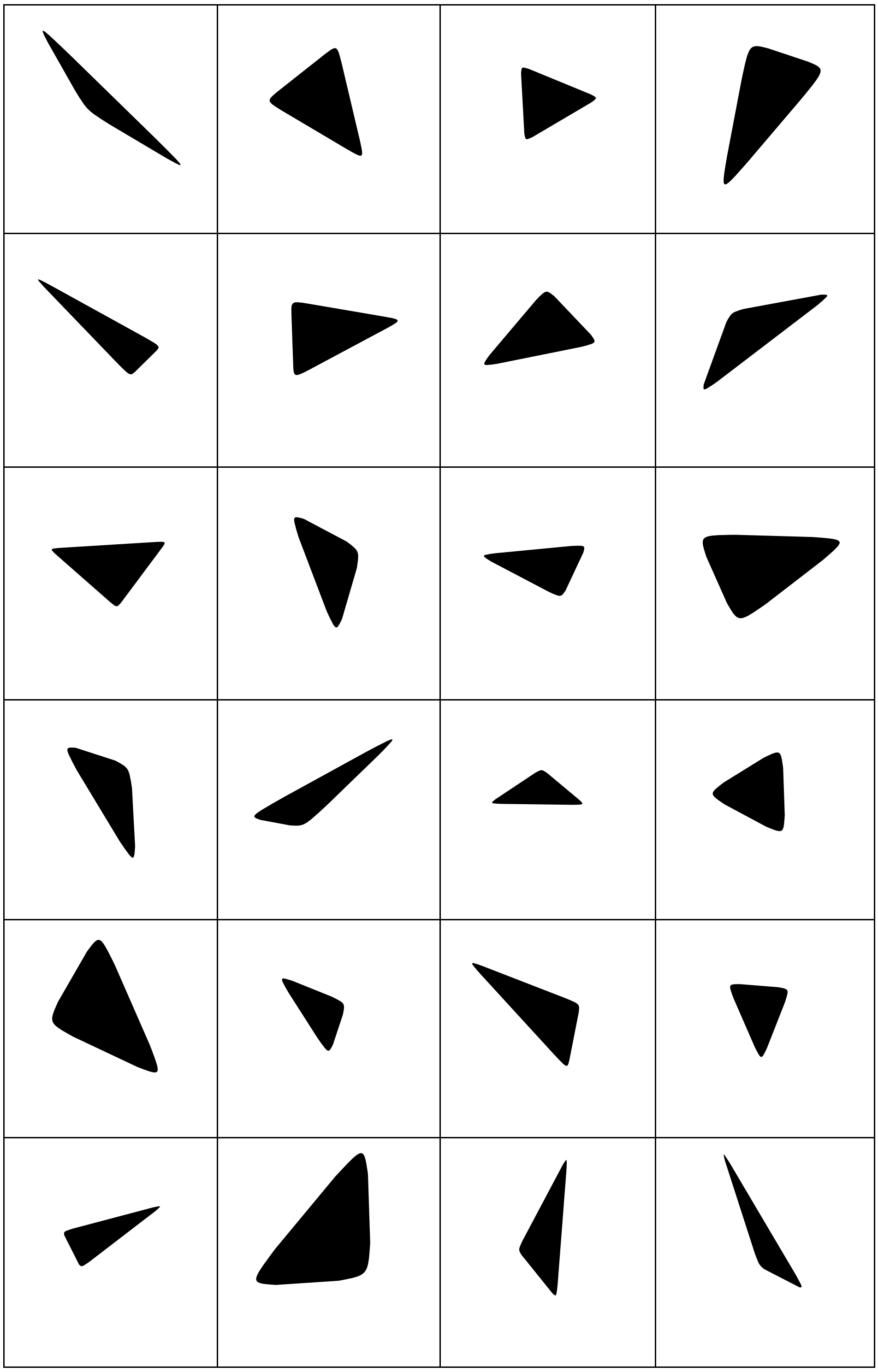
Appendix D – Experiment 1: Stimuli set D



Appendix E – Experiment 1: Stimuli set E



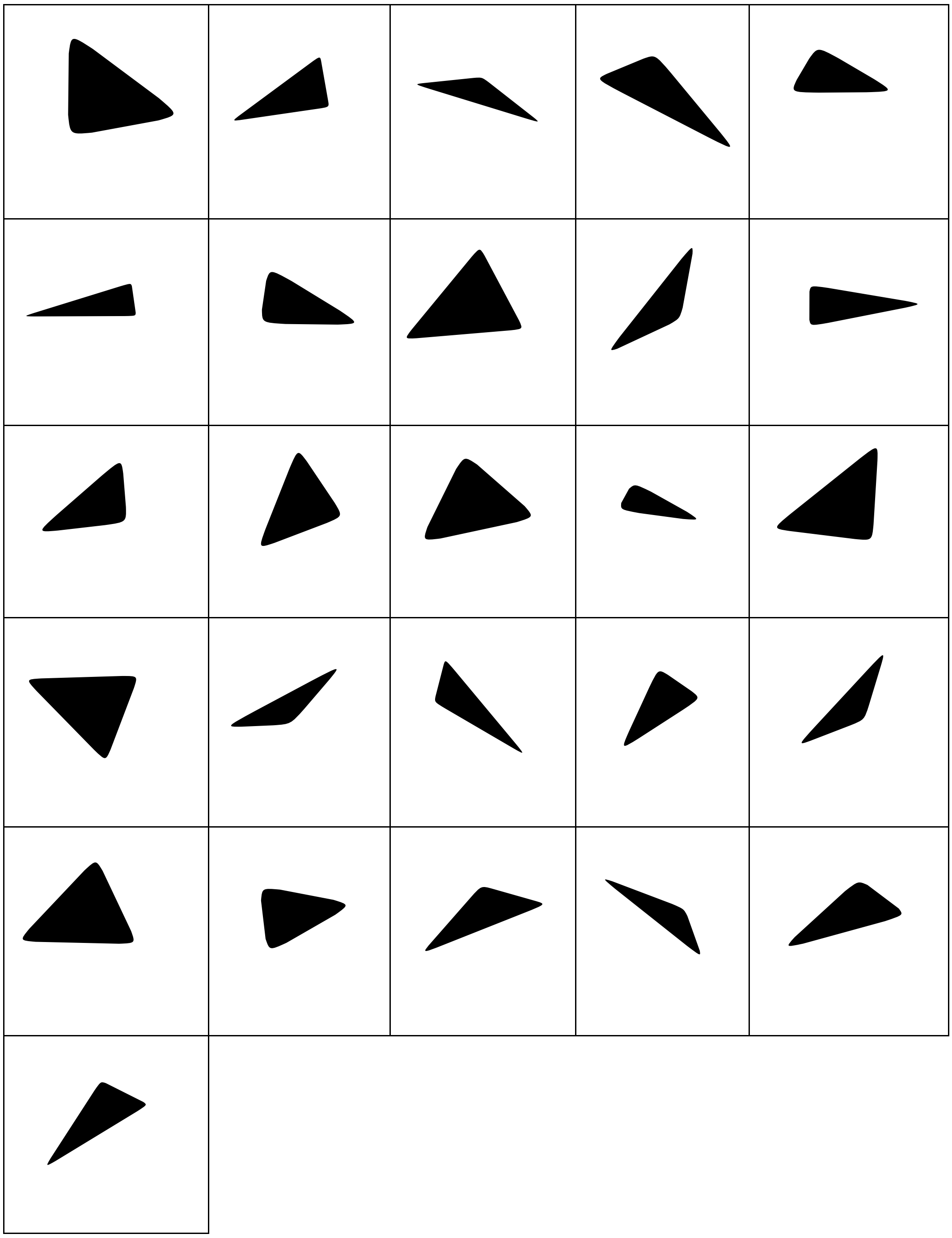
Appendix F – Experiment 1: Stimuli set F



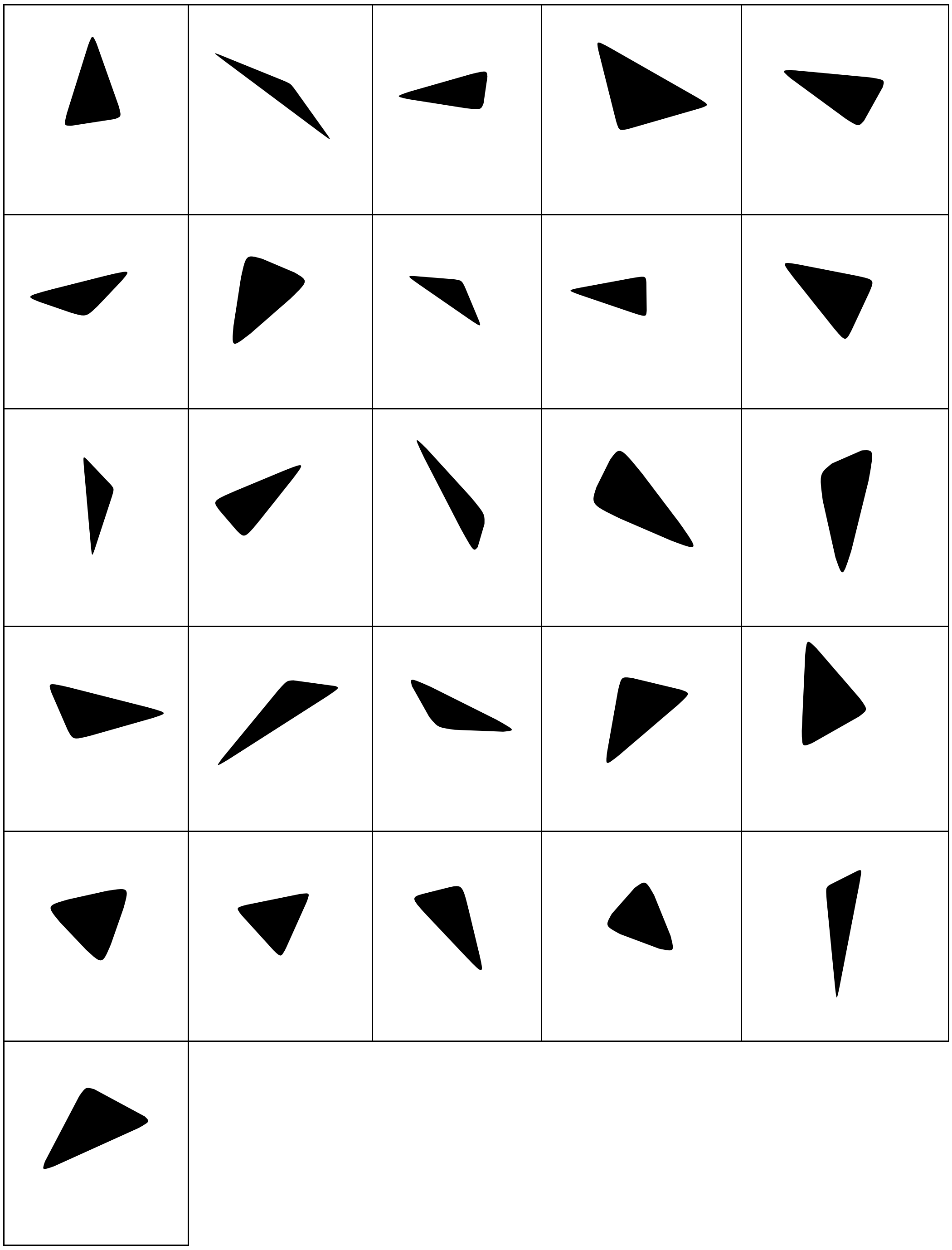
Appendix G – Experiment 1: Non-word labels

Label 1	Label 2
WEF	GIS
BIR	FOS
TUS	WEX
COH	HEK
FOZ	NER
LUD	RAS

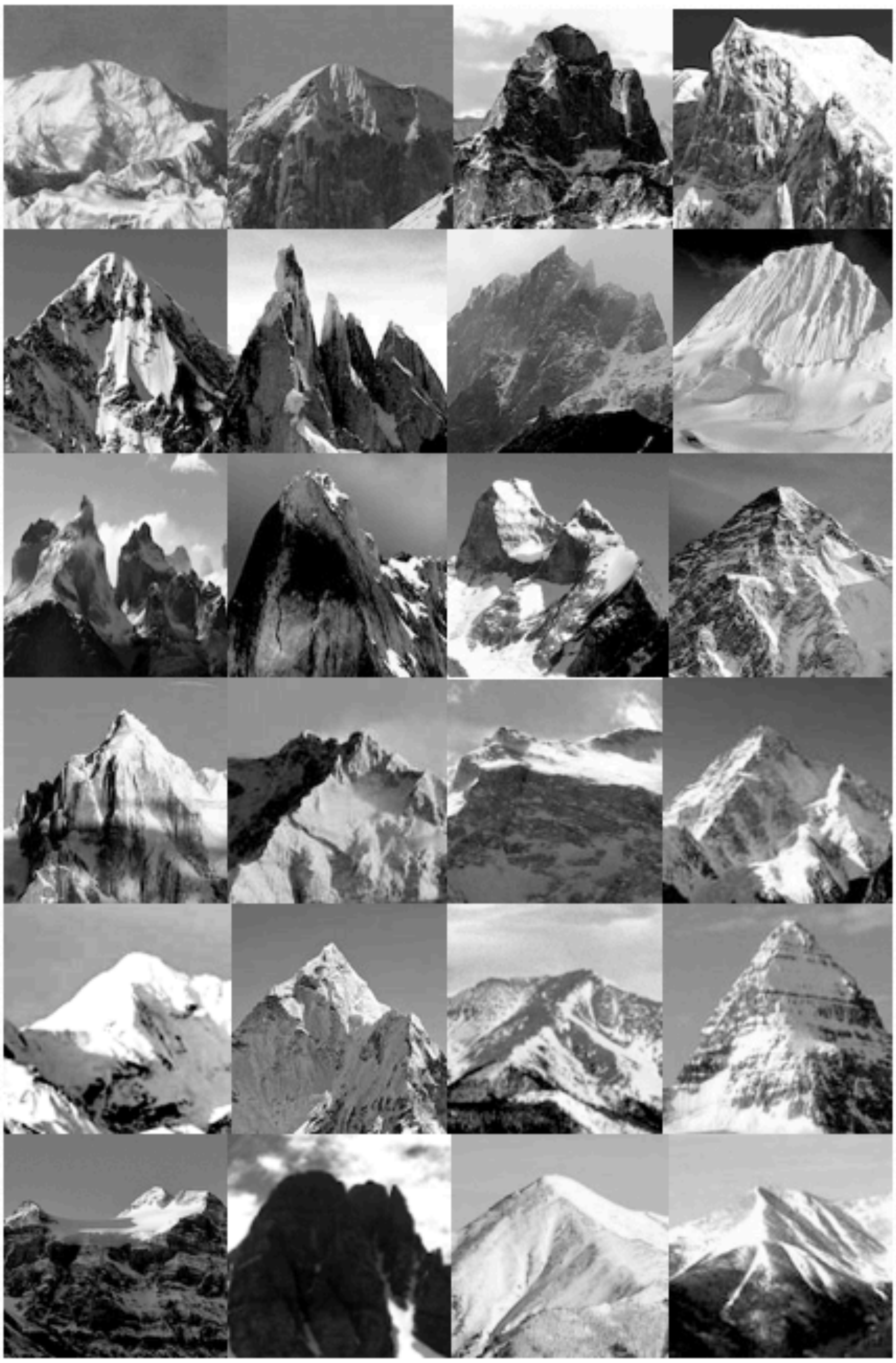
Appendix H – Experiments 2 & 3: Stimuli set A



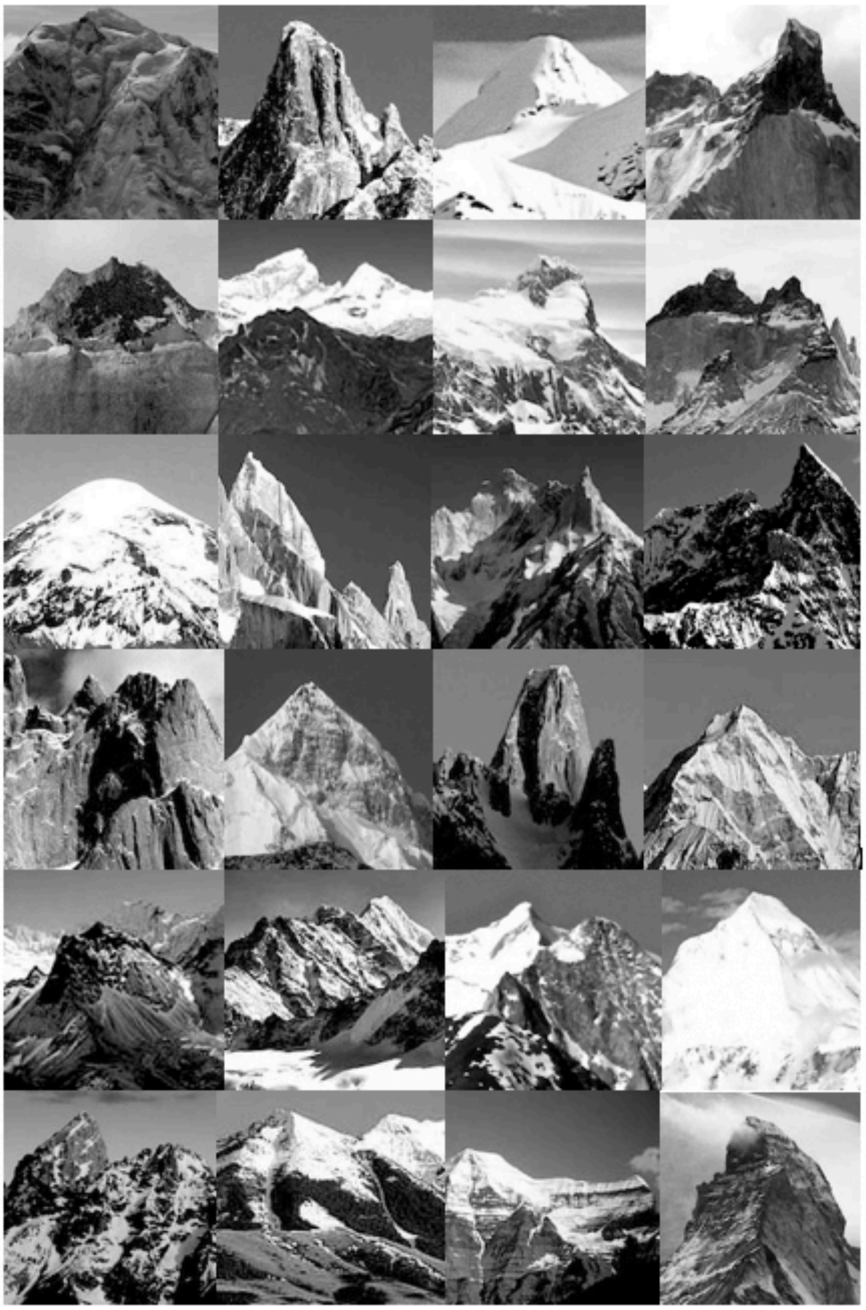
Appendix I – Experiments 2 & 3: Stimuli set B



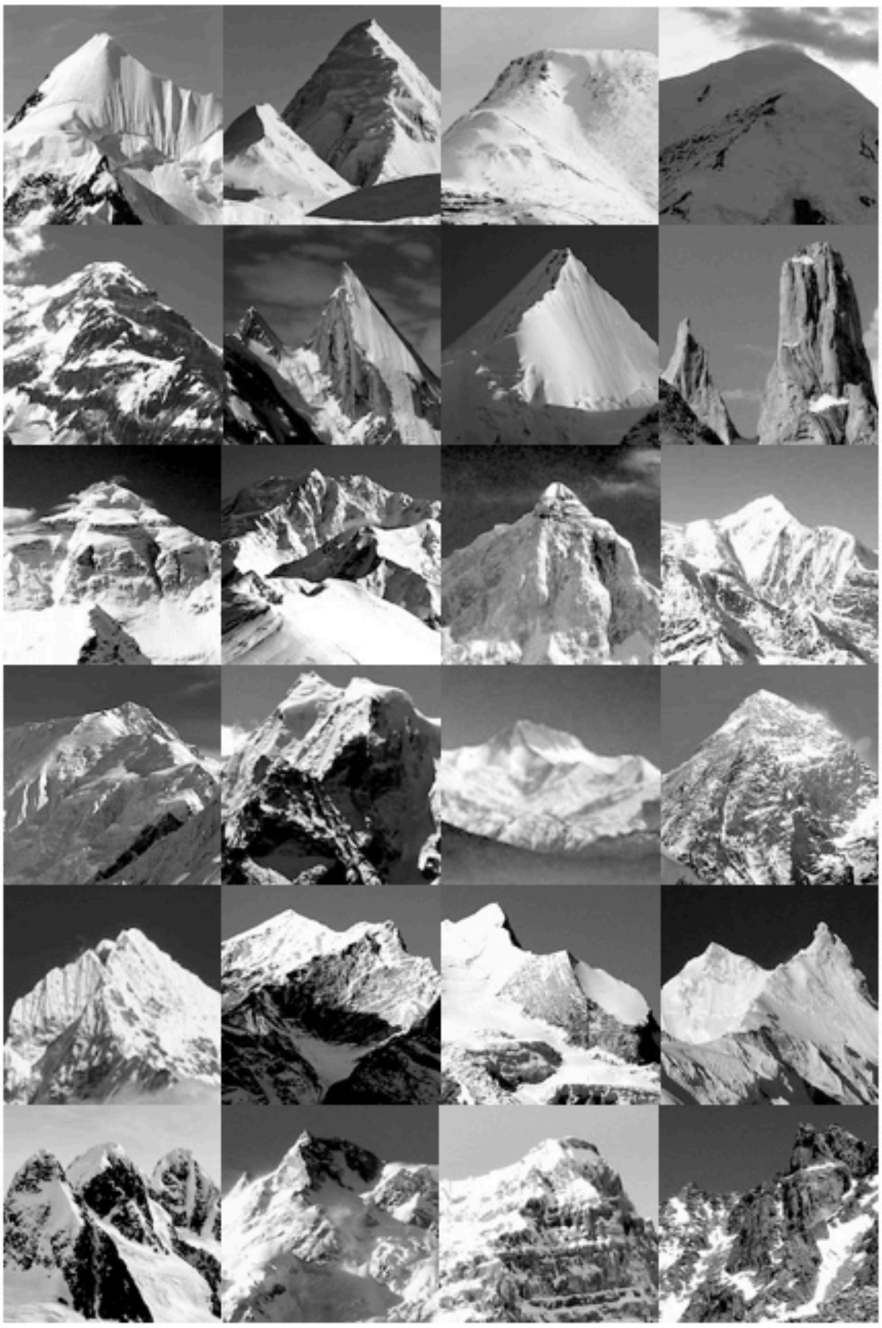
Appendix J – Experiments 4 & 5: Stimuli set A



Appendix K – Experiments 4 & 5: Stimuli set B



Appendix L – Experiments 4 & 5: Stimuli set C



Appendix M – 50 non-word label pairs

Pair no.	Label 1	Label 2	Pair no.	Label 1	Label 2	Pair no.	Label 1	Label 2
1	wex	qul	21	jik	doz	41	juz	fej
2	tij	niy	22	zum	rem	42	cuy	vap
3	qur	gul	23	zel	sov	43	zih	dor
4	fub	yuh	24	bub	mox	44	kes	voy
5	hoz	jic	25	goj	gow	45	mes	qub
6	fos	mas	26	heg	vew	46	val	mev
7	jeb	noh	27	vor	zik	47	pol	xib
8	vah	muz	28	yok	gaz	48	huz	fex
9	baj	dod	29	wix	xuq	49	kuy	las
10	pih	mul	30	fot	jaq	50	vey	jor
11	jix	fip	31	kis	dow			
12	rah	jib	32	vix	pib			
13	nal	lax	33	juy	yej			
14	fom	vaf	34	gid	luj			
15	moh	vuz	35	muv	tih			
16	tey	rew	36	xif	riy			
17	vil	gax	37	zaw	ret			
18	bir	fod	38	nex	fif			
19	fuh	mar	39	yal	bek			
20	sif	woc	40	kep	jos			

Appendix N – Experiments 6 & 7: Excerpt of LEAP-Q

Language: English

This is my (**native** **second** **third** **fourth** **fifth**) language.

(1) Age when you…

began acquiring this language:	*became fluent in this language:*	*began reading in this language:*	*became fluent reading in this language:*

(2) Please list the number of years and months you spent in each language environment:

	Years	Months
A country where this language is spoken		
A family where this language is spoken		
A school and/or working environment where this language is spoken		

(3) Please circle your *level of* ***proficiency*** in speaking, understanding, and reading in this language:

Speaking

0	1	2	3	4	5	6	7	8	9	10
None	Very low	Low	Fair	Slightly less than adequate	Adequate	Slightly more than adequate	Good	Very good	Excellent	Perfect

Understanding spoken language

0	1	2	3	4	5	6	7	8	9	10
None	Very low	Low	Fair	Slightly less than adequate	Adequate	Slightly more than adequate	Good	Very good	Excellent	Perfect

Reading

0	1	2	3	4	5	6	7	8	9	10
None	Very low	Low	Fair	Slightly less than adequate	Adequate	Slightly more than adequate	Good	Very good	Excellent	Perfect

(4) Please circle how much the following factors contributed to you learning this language:

Interacting with friends

0	1	2	3	4	5	6	7	8	9	10
Not a contributor	Minimal contributor				Moderate contributor					Most important contributor

Interacting with family

0	1	2	3	4	5	6	7	8	9	10
Not a contributor	Minimal contributor				Moderate contributor					Most important contributor

Reading

0	1	2	3	4	5	6	7	8	9	10
Not a contributor	Minimal contributor				Moderate contributor					Most important contributor

Language tapes/self-instruction

0	1	2	3	4	5	6	7	8	9	10
Not a contributor	Minimal contributor				Moderate contributor					Most important contributor

Watching TV

0	1	2	3	4	5	6	7	8	9	10
Not a contributor	Minimal contributor				Moderate contributor					Most important contributor

Listening to the radio

0	1	2	3	4	5	6	7	8	9	10
Not a contributor	Minimal contributor				Moderate contributor					Most important contributor

(5) Please circle to what extent you are currently exposed to this language in the following contexts:

Interacting with friends

0	1	2	3	4	5	6	7	8	9	10
Never	Almost Never				Half of the time					Always

Interacting with family

0	1	2	3	4	5	6	7	8	9	10
Never	Almost Never				Half of the time					Always

Watching TV

0	1	2	3	4	5	6	7	8	9	10
Never	Almost Never				Half of the time					Always

Listening to radio/music

0	1	2	3	4	5	6	7	8	9	10
Never	Almost Never				Half of the time					Always

Reading

0	1	2	3	4	5	6	7	8	9	10
Never	Almost Never				Half of the time					Always

Language-lab/self-instruction

0	1	2	3	4	5	6	7	8	9	10
Never	Almost Never				Half of the time					Always

(6) In your perception, how much of a foreign accent do you have in this language?

0	1	2	3	4	5	6	7	8	9	10
None	Almost none	Very light	Light	Some	Moderate	Considerable	Heavy	Very heavy	Extremely heavy	Pervasive

(7) Please circle how frequently others identify you as a non-native speaker based on your accent in this language:

0	1	2	3	4	5	6	7	8	9	10
Never	Almost Never				Half of the time					Always

Appendix O – Experiments 6 & 7: Stimuli set A

Appendix P – Experiments 6 & 7: Stimuli set B

Appendix Q – Experiments 6 & 7: Stimuli set C

Appendix R – Experiments 6 & 7: Stimuli set D

Appendix S – Experiments 6 & 7: Stimuli set E

References

Althaus, N., & Mareschal, D. (2014). Labels direct infants’ attention to commonalities during novel category learning. *PloS ONE*, *9*(7), e99670.

Althaus, N., & Plunkett, K. (2016). Categorization in infancy: Labelling induces a persisting focus on commonalities. *Developmental Science*, *19*(5), 770-780.

Ameel, E., Malt, B. C., Storms, G., & Van Assche, F. (2009). Semantic convergence in the bilingual lexicon. Journal of memory and language, 60(2), 270-290.

Ameel, E., Storms, G., Malt, B. C., & Sloman, S. A. (2005). How bilinguals solve the naming problem. *Journal of Memory and Language*, *53*(1), 60-80.

Arthur, B., Wemer, R., Culver, M., Lee, Y. J., & Thomas, D. (1980). The register of impersonal discourse to foreigners: Verbal adjustments to a foreign accent. In D. Larsen-Freeman (Ed.), *Discourse analysis in second language research* (pp. 111-124). Rowley, MA: Newbury House.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.

Baldwin, D. A., Markman, E. M., & Melartin, R. L. (1993). Infants' ability to draw inferences about nonobvious object properties: Evidence from exploratory play. *Child Development, 64*(3), 711-728.

Barnhart, W. R., Rivera, S., & Robinson, C. W. (2018). Effects of linguistic labels on visual attention in children and young adults. *Frontiers in Psychology*, *9*, 358.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211-227.

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. *Grounding cognition: The role of perception and action in memory, language, and thought*, 129-163.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version, 1(4).

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145-204.

Booth, A. E. (2006). Object function and categorization in infancy: Two mechanisms of facilitation. *Infancy*, *10*(2), 145-169.

Booth, A. E., & Waxman, S. (2002). Object names and object functions serve as cues to categories for infants. *Developmental Psychology*, *38*(6), 948.

Bortfeld, H., & Brennan, S. E. (1997). Use and acquisition of idiomatic expressions in referring by native and non‐native speakers. *Discourse Processes, 23*(2), 119-147.

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

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*(2), B13-B25.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1482-1493.

Casasanto, D., & Lupyan, G. (2015). All concepts are ad hoc concepts. In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 543-566). Cambridge, MA: The MIT Press.

Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences*, *3*(1), 1-15.

Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, *49*(2), 214-230.

Clark, E. V. (1988). On the logic of contrast. *Journal of Child Language*, *15*(2), 317-335.

Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.

Clark, H. H. (1985). Language use and language users. In H. H. Clark, G. Lindsey & E. Aronson (Eds.), *Handbook of social psychology (Third edition)*. New York, NY: Harper & Row.

Clark, H. H. (1991). Words, the world, and their possibilities. In G. R. Lockhead and J. R. Pomerantz (Eds.), *The perception of structure: Essays in honor of Wendell Garner.* Washington, D.C.: American Psychological Association.

Clark, H. H. (1996). Using language. 1996. Cambridge University Press: Cambridge, 952, 274-296.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on Socially Shared Cognition*, *13*(1991), 127-149.

Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language*, 332-373.

Clark, H. H. and Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshe, B. Webber & I. A. Sag (Eds.), *Elements of discourse understanding (*pp.10–63). Cambridge, UK: Cambridge University Press.

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. *Advances in Psychology*, *9*, 287-299.

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, *112*(2), 347-382.

Cook, R. G., & Smith, J. D. (2006). Stages of abstraction and exemplar memorization in pigeon category learning. *Psychological Science*, *17*(12), 1059-1067.

Costa, A., Pickering, M. J., & Sorace, A. (2008). Coherence in second language dialogue. *Language and Cognitive Processes, 23*(4), 528-556.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*(1-3), 287-314.

Dewar, K., & Xu, F. (2009). Do early nouns refer to kinds or distinct shapes? Evidence from 10-month-old infants. *Psychological Science*, *20*(2), 252-257.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences, 19*(10), 603-615.

Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition*, *143*, 93-100.

Frege, G. (1948). Sense and reference. *The Philosophical Review*, *57*(3), 209-230.

Freyd, J. J. (1983). Shareability: The social psychology of epistemology. *Cognitive Science*, *7*(3), 191-210.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181-218.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, *53*(3), 181-215.

Gelman, R. (1990). Structural constraints on cognitive development: Introduction to a special issue of Cognitive Science. *Cognitive Science, 14*, 3–10.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford, UK: Oxford University Press.

Gelman, S. A., & Davidson, N. S. (2013). Conceptual influences on category-based induction. *Cognitive Psychology*, *66*(3), 327-353.

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: the case for a developmental lexical principles framework. *Journal of Child Language*, *21*(1), 125-155.

Graham, S. A., Williams, L. D., & Huber, J. F. (1999). Preschoolers' and adults' reliance on object shape and object function for lexical extension. *Journal of Experimental Child Psychology*, *74*(2), 128-151.

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18*, 441–461.

Heine, B., & Kuteva, T. (2002). *World lexicon of grammaticalization*. Cambridge, UK: Cambridge University Press.

Heit, E. (1992). Categorization using chains of examples. *Cognitive Psychology*, *24*(3), 341-380.

Jackendoff, R. S. (2002). *Foundations of language: Brain, meaning, grammar, and evolution.* Oxford, England; Oxford University Press.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago, US: University of Chicago Press.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*, 10681–10685.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87-102.

Köhler, W. (1929). *Gestalt psychology*. New York, NY: Liveright.

Kronenfeld, D. B., Armstrong, J. D., & Wilmoth, S. (1985). Exploring the internal structure of linguistic categories: An extensionist semantic view. In J. W. D. Dougherty (Ed.). Directions in cognitive anthropology (pp. 91–113). Urbana: University of Illinois Press.

Laskowski, C. (2010). The role of language in conceptual coordination. Unpublished PhD thesis, University of Edinburgh.

Lehrer, A. (1990). Polysemy, conventionality, and the structure of the lexicon. *Cognitive Linguistics*, *1*(2), 207-246.

Levelt, W. J. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*(1-3), 1-22.

Lupyan, G. (2008). From chair to' chair': A representational shift account of object labelling effects on memory. *Journal of Experimental Psychology: General, 137*(2), 348-396.

Lupyan, G., & Casasanto, D. (2014). Meaningless words promote meaningful categorization. *Language and Cognition, 7*(2), 167-193.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking redundant labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077-1083.

Lupyan, G., & Spivey, M. J. (2010). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics*, *72*(8), 2236-2253.

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology – General*, *141*(1), 170–186.

Mackey, A., Gass, S., & McDonough, K. (2000). Do learners recognize implicit negative feedback as feedback? *Studies in Second Language Acquisition, 22*, 471-497.

Malt, B. C., & Lebkuecher, A. L. (2017). Representation and process in bilingual lexical interaction. *Bilingualism: Language and Cognition, 20*(5), 867-885.

Malt, B. C., Li, P., Pavlenko, A., Zhu, H., & Ameel, E. (2015). Bidirectional lexical discussion in late immersed Mandarin-English bilinguals. *Journal of Memory and Language, 82*, 86-104.

Malt, B. C., Sloman, S. A., & Gennari, S. P. (2003). Universality and language specificity in object naming. *Journal of Memory and Language*, *49*(1), 20-42.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Coherence and the linguistic categorization of artifacts. *Journal of Memory and Language*, *40*(2), 230-262.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940-967.

Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General, 127*(4), 331-354.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121-157.

McCawley, J. (1978). Conversational implicature and the lexicon. In P. Cole (Ed.). *Syntax and semantics 9: Pragmatics*. New York, NY: Academic Press.

McCloskey, M., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition, 6*, 462–472.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological review*, *85*(3), 207.

Munsell, A. H. (1912). A pigment color system and notation. *The American Journal of Psychology, 23*(2), 236-244.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*(3), 289-316.

Ou, L. C., Luo, M. R., Woodcock, A., & Wright, A. (2004). A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *Colour Research & Application*, *29*(3), 232-240.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280-291.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169-190.

Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, *4*(2-3), 203-228.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension?. *Trends in cognitive sciences*, *11*(3), 105-110.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329-347.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*(4), 707-727.

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception*, *22*(4), 463-475.

R Core Team (2015). R: A language and environment for statistical computing. The R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, *129*(3), 369.

Romney, A. K., Batchelder, W. H., & Weller, S. C. (1987). Recent applications of cultural consensus theory. *American Behavioral Scientist*, *31*(2), 163-177.

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive Development and Acquisition of Language* (pp. 111-144). New York: Academic Press.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192-233.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573-605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382-439.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, *47*(1), 1-24.

Schoonbaert, S., Hartsuiker, R. J., & Pickering, M. J. (2007). The representation of lexical and syntactic information in bilinguals: Evidence from syntactic priming. *Journal of Memory and Language, 56*(2), 153-171.

Silvey, C. A. (2014). Communicative emergence and cultural evolution of word meanings. Unpublished PhD thesis, University of Edinburgh.

Slobin, D. (1991). Learning to think for speaking: Native language, cognition, and rhetorical style. *Pragmatics, 1*, 7–26.

Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops?. *Cognitive Science, 34*(7), 1244-1286.

Sloutsky, V. M., Lo, Y. F., & Fisher, A. V. (2001). How much does a shared name make things similar? Linguistic labels, coherence, and the development of inductive inference. *Child Development*, *72*(6), 1695-1709.

Smith, K. (2003). Learning biases and language evolution. In S. Kirby (Ed.), *Language evolution and computation* (pp. 22–31). Proceedings of the Workshop on

Language Evolution and Computation, 15th European Summer School on Logic, Language and Information, Vienna.

Smith, J. D., Redford, J. S., & Haas, S. M. (2008). Prototype abstraction by monkeys (Macaca mulatta). *Journal of Experimental Psychology: General*, *137*(2), 390-401.

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences*, *28*(4), 469-488.

Suffill, E., Branigan, H. P., and Pickering, M. J. (2016). When the words don't matter: arbitrary labels improve categorical alignment through the anchoring of categories. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.). Proceedings of the 38th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences*, *106*(11), 4567-4570.

Thompson, B., Silvey, C., Kirby, S., & Smith, K. (2014). The effect of communication on category structure. In E. A. Cartmill, S. Roberts, H. Lyn, & H. Cornish (Eds.), *Evolution of Language: Proceedings of the 10th International Conference* (pp. 537-538). Singapore: World Scientific.

Vygotsky, L. S. (1964). Thought and language. *Annals of Dyslexia*, *14*(1), 97-98.

Waxman, S. R. (2004). Everything had a name, and each name gave birth to a new thought: Links between early word-learning and conceptual organization. In D. G. Hall & S. R. Waxman (Eds.), *From many strands: Weaving a lexicon* (pp. 295-335). Cambridge, MA: The MIT Press.

Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology*, *43*(3), 217-242.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12-to 13-month-old infants. *Cognitive Psychology*, *29*(3), 257-302.

White, A., Malt, B. C., & Storms, G. (2017). Convergence in the bilingual lexicon: A pre-registered replication of previous studies. *Frontiers in Psychology*, *7*, 2081.

Zentall, T. R., Wasserman, E. A., Lazareva, O. F., Thompson, R. K., & Rattermann, M. J. (2008). Concept learning in animals. *Comparative Cognition & Behavior Reviews, 3*, 13-45.