Title

When the Words Don't Matter: Arbitrary labels improve categorical alignmentthrough the anchoring of categories

Permalink

https://escholarship.org/uc/item/84f574sf

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 38(0)

Authors

Suffill, Ellise
Branigan, Holly
Pickering, Martin

Publication Date

2016

Peer reviewed

When the Words Don't Matter: Arbitrary labels improve categorical alignment through the anchoring of categories

Ellise Suffill (E.M.Suffill@sms.ed.ac.uk)
Department of Psychology, 7 George Square
Edinburgh, EH8 9JZ UK

Holly Branigan (Holly.Branigan@ed.ac.uk)
Department of Psychology, 7 George Square
Edinburgh, EH8 9JZ UK

Martin Pickering (Martin.Pickering@ed.ac.uk)
Department of Psychology, 7 George Square
Edinburgh, EH8 9JZ UK

Abstract

Novel labels provide feedback that may enhance categorical alignment between interlocutors. However, the nature of this feedback may not always be linguistic. Lupyan (2008) has demonstrated the effects of labels on individual categorization, and even non-word labels have seemingly produced greater consistency in sorting strategies (Lupyan & Casasanto, 2014). We extend this to alignment by demonstrating that arbitrary labels can increase sorting consistency to bring people's categories closer together, even without dialogue. Importantly, we argue that increased alignment is not always due to labeling in a linguistic sense. Results suggest that it is not the content of the non-word labels driving the alignment effects, but the very presence of the labels acting as 'anchors' for category formation. This demonstrates a more general cognitive effect of arbitrary labels on categorization.

Keywords: Alignment; Categorization; Labels; Lexical Effects.

Introduction

Shared understanding of the world is crucial for successful communication. Such *alignment* occurs through the formation of shared representations and the development of common ground (Clark & Brennan, 1991; Pickering & Garrod, 2004). Because of this, alignment is often discussed as a product of interaction. However, interaction may not always be necessary: Alignment can be considered in terms of the alignment of information states across individuals, rather than the explicit transfer of this information between them (Pickering & Garrod, 2006). For example, individuals may align without interaction if they simultaneously listen to the same speaker whilst in different rooms.

Alignment has often been studied with respect to phonological, lexical and syntactic information (Branigan, Pickering, & Cleland, 2000; Garrod & Anderson, 1987; Pardo, 2006). However, it may also include the alignment of conceptual information, and, thus, alignment can occur not only based on speakers' externally observable language, but also on their internal representations, such as alignment upon a given situation model (Zwaan & Radvansky, 1998).

We posit multidirectional feedback between several levels of linguistic information available to interlocutors. This is in contrast to feed forward accounts, such as that of Levelt, Roelofs and Meyer (1999). Instead, feedback to the conceptual level could potentially occur from any of the other, more linguistic levels. Evidence already exists for the effects of linguistic feedback on alignment across multiple levels of processing: Both word repetition and semantic relatedness can enhance syntactic alignment (Branigan, Pickering, & Cleland, 2000; Cleland & Pickering, 2003). Thus, it is possible that linguistic information such as labels could affect conceptual alignment. We will focus on the effects of word-based information on conceptual alignment, and do so through the use of categories as proxies to concepts.

According to Murphy (2002), categories act as the external application of concepts to groups of items within the world. Storing concepts as categories allows us to deal with novel objects and having exemplars related to each concept aids us in correctly categorizing novel items with other similar and previously encountered items. Thus, categories are essential to the way in which we respond to items. In language, we reinforce these categories through the application of labels that aid us in learning category boundaries (Lupyan, 2006). Thus, it seems plausible that this dependency on labels could affect categorization.

Effects of Existing Labels on Categorization

Lupyan (2008) has provided evidence for the *representational shift* account, which posits that labeling can cause a distortion of items most reliably associated with a category label, in terms of the physical features associated with that specific item. Memory was worse for items that were labeled at a basic level (e.g., chair vs. table), compared to those that were not labeled. Recall decreased more as a result of this labeling. This suggests that applying a category label to an item causes the item to become a mix of its idiosyncratic features and those typically associated with the category. Lupyan posits that this leads to greater

consistency in categories through the selection of more robust category attractors.

Lupyan and Spivey (2010) demonstrated that the effects of labeling are not limited to single items and can work across item groups. They found that hearing a redundant label leads to faster detection of visual cues placed near the 'labeled' items. For example, hearing 'chair' resulted in faster detection of a probe placed near one of multiple chairs in a display, despite participants having already been told to attend to chairs in prior instructions. Thus, labels can facilitate the process of visual perception, by affecting how individuals pay attention to groups of items, even when the labels appear redundant.

Utilizing Novel Labels

Non-word labels can be used in experiments to examine interactions between the characteristics of the labeled stimuli, and the meanings that develop for the novel labels applied to them. Using novel words has the advantage of the labels not being associated to specific entries in the mental lexicon. This gives greater potential for new label meanings to develop on the basis of the stimuli properties. Lupyan and Casasanto (2014) demonstrated that the application of meaningless, non-word labels (e.g., 'fooves' vs. 'crelches') to novel, 'alien' stimuli can facilitate the categorization process. Categorization was based upon physical characteristics, such as the smoothness versus pointedness of the aliens' heads. Results demonstrated that the novel labels worked as successfully as conventional labels (e.g., 'smooth' vs. 'pointy') at increasing sorting consistency. However, the form of the non-word labels strongly reflected sound symbolism for the physical characteristics of the items and, thus, there were expectations regarding how the labels would be applied to the stimuli (e.g., 'foove' referring to 'smooth'; cf. Köhler, 1929). Thus, it is questionable how arbitrary these labels were when placed within such a specific context. If one wanted to investigate the potential effects of more arbitrary labels, then sound symbolism should be reduced as much as possible, and stimuli could equally be made more abstract (i.e. by not having characteristics that can be readily linked to the non-word labels). Then, one could examine whether novel labels intrinsically improve consistency in categorization by developing associated meanings, if indeed labeling effects remained.

Extension to Alignment

In order to examine the nature of novel labeling effects, we developed a task that allowed us to investigate categorical alignment by utilizing non-word labels with more simplistic, abstract stimuli. For half of the experiment, participants sorted stimuli into groups with access to non-word labels, whilst in the other half of the experiment they did not have access to the non-word labels. Half of the participants were exposed to their partner's categorizations throughout, and half were not exposed. Participants were not allowed to explicitly communicate with each other in any part of the categorization task (i.e. no dialogue or gesturing).

We aimed to test whether any labeling effect remained when the non-word labels had as little influence of prior word-meaning associations as possible and when they could not be consistently linked to the stimuli via sound symbolism. Specifically, we asked whether the effect was due to the labels acquiring meaning, or whether it was due to a less linguistic process by which labeling increased consistency in participants' sorting strategies without acquiring meaning.

We also examined whether the effects of labeling were affected by exposure to one another's sorting strategies. The Exposed condition tested for the effect of labels on the alignment of categories through participants seeing their partner's categories at intervals throughout the task. The Non-exposed condition focused on whether alignment would be higher when categorization occurred alongside access to labels, regardless of the fact that participants did not have exposure to each other's categories (i.e., if there would be greater consistency between individuals when they had access to labels, compared to when they did not have labels). If the non-word labels increased sorting consistency by developing an associated meaning, then it is possible that this meaning could be communicated across the individuals within a pair, given the shared nature of words and associated meanings (Laskowski, 2010). This could lead to differences in alignment for the Exposed versus Non-exposed conditions, as information transfer as an effect on alignment could only occur in the Exposed setting.

The items to be categorized were randomly generated, triangular shapes (Laskowski, 2010). The non-word labels were provided in order to help participants divide up the stimuli into more discrete categories. The dependent variable of the study was how aligned participants were in their assignment of items to categories i.e., the extent to which their categorizations overlapped.

Method

Participants

64 British participants (45 female) formed 32 experimental pairs. Ages ranged from 18 to 27 years, with an average of 19.86 years. All participants identified themselves as native, monolingual speakers of English. Written consent was obtained prior to testing. The University of Edinburgh's Ethics Committee approved this study.

Stimuli

Perceptually morphed, triangular shapes (from Laskowski, 2010) were used as experimental items (see Figure 1). Shapes were morphed across the dimensions of size, shape, angles and pointedness. Items were printed on 3x3 inch cards. There were two sets of triangles (Set A and B), each comprising 26 items.

The non-word labels were "TEB" (/tɛb/) and "DUP" (/dʌp/). These were presented as printed labels alongside the

stimuli sets in the With-labels condition, and were absent in the No-labels condition. The labels were constructed to be similar in linguistic qualities, so as to reduce the possible effects of sound symbolism. That is, both /t/ and /d/ are alveolar plosives, /ɛ/ and /ʌ/ are open-mid, unrounded vowels (although they vary slightly in position), and /b/ and /p/ are both bilabial plosives.

Figure 1: Items 1-6 from Set A as examples of stimuli.

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 throughout the study. In the Non-exposed conditions, participants never saw the other person's categories. For Labels, participants who had access to non-word labels (With-labels) in one block (e.g., trials 1–5) would not have labels (No-labels) in the second block (e.g., trials 6–10), and vice versa. We counterbalanced in which block participants had access to labels and which set (A or B) they sorted first. 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 the intervals so that participants could see each other as well as their item groups. In the Non-exposed condition, the barrier obscured the view of the other participant and their item groups for the entirety of the experiment.

Participants were instructed to categorize the stimuli into two groups using a minimum group size of 9 and a maximum group size of 17, but could use any number between this (e.g. 12 and 12, or 16 and 10, etc.). This is because limiting group sizes to a specific number could have led to more unnatural category formation, by forcing participants to place items under categories that they would not otherwise have chosen to.

Procedure

The experimental set up and procedure was as follows, with each experiment comprising 10 rounds (see Figure 2):

1. Participants were given the instructions; "Sort the triangles into two groups in a way that would make sense to you, as well as to another person".

2. Upon receiving the labels participants were told; "You have these two labels to place upon your groups. Place one label on one group each. You choose how to use them. You can move them across groups between rounds if you wish"

3. Participants faced each other across a table with the barrier in place.

4. Participants sorted one set of A or B into two groups, with a time limit of 2 ½ minutes. All triangles had to be assigned to a group.

5. There was a 30-second interval between sets. In the Exposed conditions, participants saw their partner's groups (plus labels, if they were on the Labels block of the experiment); in the Non-exposed condition participants did not see their partner's sorting strategies.

6. The barrier was replaced (for Exposed conditions) and participants sorted the next 4 sets.

7. After trial 5 (end of block one), participants took a short break from the task in which they filled in demographic information.

8. In the second block, whichever set was not used in block one was sorted over trials 6–10.

9. Participants completed a post-test questionnaire regarding their sorting strategies and whether they applied meaning to the labels.
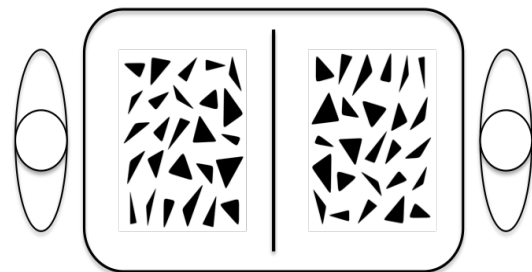

Figure 2: Two participants faced each other across a table with the central barrier in place.

Results

Calculating Alignment Scores

Scoring reflected how similarly the two participants in a pair split items across groups. For each set, a participant would finish with two groups of items. Each item was identifiable to the coder by number. If both Participants A and B placed items 1–8 in 'group 1' and items 18–26 in 'group 2', this would result in an alignment score of 17 for the round. We then reversed the assignment of groups for Participant A and determined the alignment score. We selected the higher of these two scores (as assignment of group number was arbitrary).

Descriptive Statistics

The average alignment scores for each Condition were numerically higher in With-labels blocks relative to No-labels blocks (see Table 1).

Table 1: Mean alignment scores (*SD*) by Block for each Condition of Exposure by Labels.

Condition	Mean Alignment	
	Block1	Block2
Ex+Labels1st	20.15 (4.00)	19.05 (3.82)
Ex+Labels2nd	18.6 (4.30)	19.85 (2.93)
Non-Ex+Labels1st	18.63 (3.74)	17.4 (2.76)
Non-Ex+Labels2nd	18.48 (3.27)	18.6 (3.47)

Linear Mixed Effects (LMER) Analysis

In order to test for the effects of Labels and Exposure on alignment, data were analyzed in R 3.2.1 (R Core Team, 2015) with a linear mixed-modeling approach via the lme4 package, version 1.1-8 (Bates, Maechler, Bolker & Walker, 2013). This approach was chosen as it allowed us to account for random variance due to differences between participant pairs. Under this approach, the threshold for statistical significance was set at $|t| > 2$. 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 variable Labels, with Pair (participant pairs) also entered as a random effect. The initial model included Set (A or B), Labels (With-labels or No-labels), Exposure (Exposed or Non-exposed) and Order (Labels 1st or Labels 2nd). Removing Order as a fixed effect did not significantly reduce model fit ($X^2(1)=0.61$, $p=.44$). Removing Exposure as a fixed effect also did not significantly reduce model fit ($X^2(1)=2.07$, $p=.15$). However, removing Labels as a fixed effect did significantly reduce model fit ($X^2(1)=6.13$, $p=.01$), as did removing Set ($X^2(1)=7.51$, $p=.01$). Including the interaction between Set and Labels did not contribute significantly to model fit ($X^2(1)=2.46$, $p=.12$). The best-fit model therefore included the main effects of Set and Labels as fixed effects (see Table 2). With-labels produced significantly higher alignment scores than No-labels ($\beta=0.93$, SE=0.37, $t=2.52$). Thus, there was a significant effect of Labels, regardless of Exposure. Set A also produced significantly higher alignment scores than Set B ($\beta=1.04$, SE=0.37, $t=2.83$).

Table 2: Beta, standard errors and t-values for fixed effects on alignment score. Model fit by REML.

Fixed Effects	β	S.E.	t
Intercept	-0.98	0.50	-1.96
Set(A)	1.04	0.37	2.83
Labels(With)	0.93	0.37	2.52

Table 3: Variance and residual for random effects. Model fit by REML.

Random Effects		
Pair	Intercept	5.383
	Labels(With)	1.252
Residual		7.67

No. of Observations = 320.

Labels-specific Analysis

A labels-specific analysis was conducted using LMER to assess whether having labels matched across groups led to significantly higher alignment scores. Matched status was dependent on whether the non-word labels matched across the groups used to calculate a pair's alignment scores: that is, whether Participant A placed specific items under TEB and others under DUP in the same manner as Participant B. For example, if Participant A's group 1 was compared to Participant B's group 1 in order to calculate their alignment score, matched status would depend on whether both participants placed the same label upon that group. There were 160 data points overall for the With-labels condition, with 4 similarly sized groups in terms of data points across Match and Exposure conditions (see Table 4).

Table 4: No. of observations and mean alignment (*SD*) for Matched vs. Unmatched labels for Exposed and Non-exposed Conditions

Exposure	Match	*N*	Mean (*SD*)
Exposed	Matched	42	20.76 (2.99)
	Unmatched	36	19.26 (3.87)
Non-exposed	Matched	38	18.81 (3.83)
	Unmatched	44	18.48 (3.51)

Models included random slopes and intercepts for Set, Exposure and Match (Matched or Unmatched), with Pair (participant pairs) entered as a random effect. The initial model included Set, Exposure, Match and Order as fixed effects. Removing Order did not significantly reduce model fit ($X^2(1)=1.02$, $p=.31$). Removing Match also did not significantly reduce model fit ($X^2(1)= 0.48$, $p=.49$). Removing Exposure significantly reduced model fit ($X^2(1)= 4.23$, $p=.04$), as did removing Set ($X^2(1)= 3.91$, $p=.04$). Including the interaction between Set and Exposure did not significantly increase model fit ($X^2(1)=1.00$, $p=.32$). Therefore, the model of best fit included Set and Exposure as a fixed effects (see Table 5). The Exposed condition produced significantly higher scores than the Non-exposed condition ($\beta=2.07$, SE=0.79, $t=2.61$). Set B produced significantly lower scores than Set A ($\beta=-1.74$, SE=0.79, $t=-2.19$).

Table 5: Beta, standard errors and t-values for fixed effects on alignment score. Model fit by REML.

Fixed Effects	β	S.E.	t
Intercept	19.16	0.61	31.28
Set(B)	-1.74	0.79	-2.19
Exposure(Expo)	2.07	0.79	2.61

Table 6: Variance and residual for random effects. Model fit by REML.

Random Effects		
Pair	Intercept	0.74
	Set(B)	4.33
	Exposure(Expo)	9.48
	Match(Match)	2.54
Residual		7.81

No. of Observations = 160.

Qualitative Analysis

Data from post-test questionnaires were analyzed to check for patterns in assigned label meanings. In 32 instances (50% of participants), one or both of the labels were reported to hold no significance to the individual. There was a great deal of variation in the meanings that the remaining participants assigned to the labels, as reflected by the 34 different terms that they used to refer to them. These ranged across references to stimuli size, length, angle size, roundness of corners and even number of items.

Sound symbolism effects associated with the slight differences in vowel position for TEB and DUP appear highly unlikely to explain the effects: There were only 3 instances in which both participants within a pair reported assigning meanings to the label that could be considered semantically similar. For example, one participant associated DUP with 'Fat' and TEB with 'Thin', whereas their partner associated DUP with 'Heavy' and TEB with 'Light'. This high variability in meanings and low occurrence of matches in label associations is consistent with our finding that label content did not significantly contribute to the increased alignment scores across participants in a pair, at least on an explicit level.

Discussion

Pairs of participants were more likely to categorize abstract stimuli in the same way when they had access to non-word labels, than when they did not. This is consistent with Lupyan and Casasanto's (2014) proposals, in that it suggests that arbitrary, non-word labels may gain significance in meaning and, through this, focus the categorization process to create more consistent sorting patterns. In the current study, the labeling effect held regardless of exposure (i.e., whether participants saw each other's sorting strategies, or not), which suggests labels may increase consistency in sorting across individuals, so that people's categorizations are more aligned without interaction. However, examining data only from trials in which participants did have access to labels does suggest an effect of exposure. It appears that whilst arbitrary labels can improve alignment without the need for interaction, having exposure to a partner's sorting strategies can build upon this labeling effect.

Exposure Effects in Alignment

Language transmission studies have demonstrated how novel language systems for open-ended meaning spaces can begin to evolve categorical structure with enough exposure across generations of participants, as sorting patterns are transmitted from one individual to the next, and as a function of how much iteration occurs (Kirby, Cornish & Smith, 2008; Carr, Cornish & Kirby, 2015). This categorical structure appears to develop on the basis of cognitive biases for physical features such as item shape and size.

As such, these processes could occur within the current paradigm, but with information being repeatedly passed between a pair of individuals, rather than being transmitted down from one generation to the next. For internal consistency in the non-exposed condition, iteration would reinforce sorting strategies through repeated exposure to items, in order to increase overall consistency within and across individuals (and this would be facilitated by the presence of non-word labels). As for alignment in the exposed condition, a similar scenario could develop, but with transmission occurring at the interval in which participants transfer information regarding sorting strategies via exposure.

Labeling vs. Anchoring

We found that, whereas alignment scores were higher in the presence of non-word labels, alignment was not affected by specifically which items the labels were applied to. This was corroborated by the qualitative analysis of label meanings reported by participants in the post-test questionnaire. Although it is possible that implicit labeling effects could have occurred, it does not seem that any consistent meaning was tied to TEB or DUP in a way that would enhance alignment.

Instead, the labels may have led to higher categorical alignment through a form of 'anchoring', by providing a point of reference for each category, but without explicitly linking to that category through an associated meaning. In a sense, anchoring can be thought of in terms of Kirsh's (1995) markers, as opposed to perfect reminders. That is, the labels may act as markers that indicate to the individual that there is an important distinction to be made between the items (the two categories into which they are to be sorted), without telling us anything specific about the nature of that distinction. Existing labels (or novel labels that have gained an associated meaning) act as perfect reminders, in that they can communicate something important about the category (and category items) that they are applied to. Thus, the anchoring effect demonstrated here could not be readily referred to as linguistic, but is instead something seemingly more general, perhaps linking to other cognitive functions

commonly considered as external to the linguistic domain. Specific, cognitive functions that language is argued to support include memory, attention and simplification of tasks, as well as the development of representations (Clark, 1998). Given this, the effect could stem from having the label there as a "word", rather than it mattering specifically what that word is.

Conclusion

The effect of labels on categorical alignment is seemingly not always due to a linguistic process of associating concepts to words as linguistic units. In the current study, the labels did not appear to lead to higher categorical consistency and alignment through linguistic labeling effects, but through a more general cognitive process of anchoring, or marking out the distinction between categories. This finding has important implications for studies investigating what may appear to be 'labeling' effects. The non-word labels led to higher alignment even in the absence of exposure between pairs, demonstrating that alignment does not rely exclusively on interaction and sometimes can occur through increased consistency across non-interacting individuals. However, it does appear that exposure to a partner's sorting strategies can build upon the effects of labeling on alignment. Future studies should aim to discover the generalizability of this anchoring effect, by introducing new forms of potential markers to the categorization process, as well as by investigating at what point the markers could potentially gain communicative meaning through interaction.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and S4. *R package version*, *1*(4).

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

Carr, J. W., Cornish, H., & Kirby, S. (2014, April). The cumulative cultural evolution of category structure in an open-ended meaning space. Paper presented at the 10th International Conference on the Evolution of Language, Vienna, Austria.

Clark, A. (1998). Magic words: how language augments human computation. In P. Carruthers and J. Boucher, eds, *Language and thought: interdisciplinary themes* (pp.162–183). Cambridge, MA: Cambridge University Press.

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

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.

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*(31), 10681-10686.

Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence*, *73*(1), 31-68.

Köhler, W. (1929). *Gestalt psychology*. Oxford, England: Liveright.

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

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(01), 1-38.

Lupyan, G. (2006). Labels facilitate learning of novel categories. In A. Cangelosi, A.D.M. Smith, & K. Smith (Eds.), *The Sixth International Conference on the Evolution of Language* (pp. 190–197). Singapore: World Scientific.

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

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

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

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

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382-2393.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(02), 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.

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

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162.