Individual Differences in Concept Dominance

Willem Jacob Louter (568358wl@eur.nl)
Department of Psychology, Education and Child Studies,
Erasmus University Rotterdam, Post Box 1738 3000 DR Rotterdam, The Netherlands

Sam Jabbari (sam.jabbari98@gmail.com)
Department of Psychology, Education and Child Studies,
Erasmus University Rotterdam, Post Box 1738 3000 DR Rotterdam, The Netherlands

Sandra Blythin Constantinou Juhasz (constantinou.sandra@gmail.com)
Department of Psychology, Education and Child Studies,
Erasmus University Rotterdam, Post Box 1738 3000 DR Rotterdam, The Netherlands

Steven Verheyen (verheyen@essb.eur.nl)
Department of Psychology, Education and Child Studies,
Erasmus University Rotterdam, Post Box 1738 3000 DR Rotterdam, The Netherlands

Abstract

The literature on conceptual combination has thus far been limited to research at the aggregate level investigating adjective-noun and noun-noun combinations. One well-established phenomenon within this literature is that of concept dominance (Hampton, 1988), which is the finding that the relative contribution of constituent concepts (for example *sport* or *game*) to their conjunction (*sport that is also a game*) is often very unequal. This exploratory study investigated individual differences in how people understand adjective-adjective-noun combinations, such as *long blue coat*. Participants rated images of coats varying along the perceptual dimensions of length and color for typicality in two different conjunctions, namely *long blue coat* and *long purple coat*. We used multidimensional scaling (MDS) to construct an aggregate coat space from similarity data collected with the Spatial Arrangement Method (SpAM). Using external unfolding, we modeled participants' typicality judgements by representing their individual typicality data as vectors within the aggregate MDS space, such that orthogonal projections from the coats onto the vectors represent their perceived typicality in the conjunctions. We did not find strong evidence for concept dominance at the aggregate level; however, we did find evidence for concept dominance at the individual level, with marked individual differences in the extent of dominance and which dimension was dominant. The validation of external unfolding for research into conceptual combination comes with new research possibilities, several of which are proposed.

Keywords: concept combination; conjunctions; external unfolding; concept dominance; individual differences

How people combine ill-defined semantic concepts from everyday language (for example *sport* or *game*) in order to understand conjunctions of these concepts (for example *a sport that is also a game*) has been a long-standing topic of psychological research. These concepts are generally vague (McCloskey & Glucksberg, 1978; Verheyen, Hampton & Storms, 2010), as there are no clear boundaries that distinguish between for example an activity that is a sport and an activity that is not (e.g., Is darts a sport or not?). Yet, people are still able to understand the conjunctions of these vague concepts in meaningful ways. The processes underlying conceptual combination are still not fully understood and have important implications for general theories on how people conceptualize (Hampton, 1997). A popular way of studying conceptual combination has been to investigate how an object's typicality within constituent concepts relates to its typicality within the conjunction of these concepts.

Using regression equations, Hampton (1988) found that the weighted average of constituent typicality could reliably predict conjunction typicality. An unexpected finding, however, was that for many concept pairs the relative contribution (expressed through the regression weight) of the constituents' typicality to conjunction typicality was very unequal. For example, in predicting conjunction typicality for *sports that are also games* and *games that are also sports*, typicality within the concept *sport* was given almost twice the amount of weight as was given to typicality within the concept *game* (Hampton, 1987, 1988). Hampton termed this phenomenon concept dominance. Concept dominance within conjunctions is well-established and oft replicated (Hampton, 1988, 1997; Storms, De Boeck, Van Mechelen, & Geeraerts, 1993; Storms, De Boeck, Van Mechelen, & Ruts, 1996).

Research into concept dominance has so far been limited to combinations of adjective-noun pairings (Hampton, 1996), noun-noun pairings (Hampton, 1988), and relative clause descriptions (*A that is also B*) of noun-noun pairings (Hampton, 1987, 1988; Storms et al., 1993, 1996), with Storms, Ruts and Vandenbroucke (1998) showing that the phenomenon generalizes across these syntactic forms. We propose that the phenomenon's generalization to adjective-adjective-noun combinations (for example, *small red cup*) merits investigation too because these combinations occur very often in everyday language. They have the additional advantage that they can easily be studied experimentally through the presentation of concrete visual stimuli that

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

systematically vary along both adjectives (see Figure 1 for examples), whereas much of the work on noun pairings has used abstract verbal materials that are less under the experimenters' control.

To date, only one study has investigated adjective-adjective-noun combinations. Hampton (1996) used images of faces that varied along dimensions of happiness (from happy to sad) and either intelligence (from intelligent to stupid), or age (from child to adult), thus investigating eight different conjunctions (for example, *happy adult face*). Participants rated the images for typicality in both constituents and the conjunction. No conclusion could be drawn regarding concept dominance however, as the interpretation of the regression analysis was compromised due to differences between constituents in the reliability of typicality ratings. In line with Hampton's (1996) approach, our study will make use of a set of 25 images[1] depicting coats that vary along two perceptual dimensions, namely length (from long to short) and color (from blue to purple). Participants will rate the typicality of these images for two different conjunctions, namely *long blue coat*, and *long purple coat.*

Another limitation of the reviewed studies is that the processes involved in combining constituent concepts into conjunctions were analyzed at the aggregate level. Although the research summarized thus far often reports excellent reliability for typicality ratings, this reliability has been estimated by measuring the stability of the mean of these typicality ratings and does not actually measure between subject agreement (Barsalou, 1987). This means individual differences in the typicality ratings cannot be ruled out, as correlating the mean typicality ratings of for example the split halves will result in high reliability if the range of the individual differences is consistent across the sample or if the sample size is sufficiently large. Indeed, as shown by Barsalou (1987), the average of the interindividual correlations gives better insight into individual differences in typicality ratings and is often much weaker than the reliability would seem to suggest. Furthermore, Barsalou (1989) argues that in averaging across participants, effects from idiosyncratic sources of information on concepts are lost, and only effects of shared information on the concept remain. This also means that it is possible that the averaged typicality data do not represent individuals' actual typicality judgements. When the average is not representative of the behavioral patterns of any of the individuals within a sample, one might draw wrong conclusions regarding the behavioral patterns within the population (Martin & Caramazza, 1980; Verheyen & Storms, 2007). For these reasons, we argue that a focus on individual differences might provide further insight into some of the processes that are involved in conceptual combination. Therefore, the present study will focus on individual differences in typicality within conjunctions of adjectives.

To model how participants perceive typicality within conjunctions of adjectives, we will make use of a statistical procedure called external unfolding. External unfolding can be understood as a special kind of multidimensional scaling (MDS; Borg & Groenen, 2005). In MDS, (dis)similarity among objects is spatially represented in such a way that the distance between objects reflects how (dis)similar they are (i.e., objects that are more similar are positioned close to each other, and objects that are more dissimilar are positioned further apart; Borg & Groenen, 2005; Torgerson, 1952). For example, if one were to take the distances between all European capitals (in km) as the input for an MDS procedure, the output would be a scaled map, with each of the capitals represented in the MDS space at appropriate relative distances from one another. With the use of external unfolding, individual participants can be represented within such an MDS space of object (dis)similarity, based on their individual preferences (or in our case, typicality ratings) for each of the objects. Using the VIPSCAL algorithm (Van Deun, Groenen & Delbeke, 2005), we will estimate a vector model (Tucker, 1960) which represents participants within the MDS space as directed lines (i.e., vectors), with orthogonal projections of objects onto the vector representing the rank order of typicality so that the higher an object projects onto the vector, the more typical the object was rated by the participant (for a similar idea using mean typicality ratings, see Voorspoels, Vanpaemel & Storms, 2011). Additionally, we will also represent the average typicality data with such a vector, which will allow us to establish the representativeness of the aggregate representation for the individual typicality data.

In order to obtain the necessary (dis)similarity data for our analyses, participants will perform a spatial arrangement task. The Spatial Arrangement Method (SpAM) was developed by Goldstone (1994) as an efficient alternative to obtaining similarity ratings by way of pairwise comparisons. In SpAM, many stimuli are presented at once on a computer screen, and participants spatially arrange the stimuli in such a way that the distance between them represents how (dis)similar they find them. Hout, Goldinger and Ferguson (2013) found that SpAM is an efficient way to obtain similarity data which still leads to high quality MDS spaces.[2]

In sum, the present exploratory study aims to use external unfolding to investigate whether evidence for concept dominance can be found for adjective-adjective-noun combinations, with a specific focus on individual differences between participants.

[1] The study also included a set of 25 images depicting glasses varying in size and shade, and a set of 25 images depicting tomatoes varying in size and color. Since the results obtained with these different stimuli were comparable, we will only discuss coats here because of page constraints.

[2] Verheyen, Voorspoels, Vanpaemel, & Storms (2016) provide several important caveats but they do not apply to the current study due to the two-dimensional visual nature of the stimuli and the fact that an aggregate space is used (see below and also Verheyen, White, & Storms, 2022).

Method

Participants

Our sample consisted of 66 undergraduate students of whom 54 identified as female and 12 identified as male. Their age ranged from 17 to 28 years old (*M* = 19.86, *SD* = 1.92). One participant indicated to have a visual impairment (tritanomaly) but as we were focused on individual differences, we decided not to exclude this participant. Participants were recruited using the central registry of Erasmus University Rotterdam and were rewarded with partial course credit.

Materials

The experiment required a computer with the E-Prime 3.0 software (Psychology Software Tools) installed and access to the January 2023 version of the online survey software Qualtrics (Qualtrics, Provo, UT). Monitors had a resolution of 1920x1200, with a 24-inch screen size.

Stimuli were taken from a not-yet published pre-registered study (Verheyen, Egré, Hampton, Scerrati, & Urbonaviciute 2022) and consisted of 25 images depicting coats varying in size (from long to short) and color (from blue to purple). This set of images was validated by Verheyen and colleagues for studying the conjunctions of *long blue coat* and *long purple coat* by measuring assent within the constituents. From a larger set of images, they selected stimuli for which either 0%, 25%, 50%, 75%, or 100% of participants rated the image as a member of each constituent. For example, in the case of the longest and most blue coat, 100% of participants rated the image as a long coat, and 100% of participants rated it as a blue coat. To illustrate the dimensions along which the stimuli varied, extremes of length and color are presented in Figure 1. Stimuli names refer to their relative position within the two dimensions, with L standing for length, and C standing for color. For example, L5 = longest, and C1 = most blue.

L5C1 L1C5

Figure 1: Extremes of length and color

Procedure

Participants were tested individually in soundproofed cubicles and each participant provided informed consent.

The spatial arrangement task began with the 25 images ordered randomly in a 5x5 grid. Participants were instructed to organize the images so that the difference between them reflected their similarity (i.e., more similar images placed closer together, and more different images placed further apart). It is important to note that the dimensions along which the images varied were not mentioned in the instructions.

The typicality survey followed the spatial arrangement task. This order was chosen to prevent the possibility that participants' performance on the spatial arrangement task might be influenced by already having been presented with the underlying dimensions (for example *long blue coat*) during the typicality survey. In the typicality survey, participants were presented with instructions taken from Rosch and Mervis (1975) and asked to rate the typicality of the coat images for the conjunctions of *long blue coat* and *long purple coat*, with the order of conjunctions randomized for each participant. For both conjunctions, all 25 images were individually presented in random order in the middle of the screen, accompanied with a 7-point Likert scale and the following question: "How good an example of the category LONG BLUE/PURPLE COAT is this image?". After all images had been rated for both conjunctions, participants were thanked for their participation.

Results

Multidimensional Scaling

The reliability of the dissimilarity data (distance in pixels) obtained during the spatial arrangement task was estimated by applying the Spearman Brown formula to the split-half correlation, resulting in a reliability of .95. This high reliability suggested that it would be safe to assume a shared representation of the stimuli among the participants.

To see whether participants' spatial arrangements captured the underlying dimensions of the stimuli, multidimensional scaling (MDS) was performed on the dissimilarity data. Using the SMACOF package (de Leeuw & Mair, 2009; Mair, Groenen & de Leeuw, 2022) in R 4.3.0. (R Core Team, 2023), two-dimensional ordinal MDS was fitted to the dissimilarity data (averaged across participants). The ordinal level was chosen so as to impose the fewest number of restrictions on the data, and a two-dimensional space was chosen as we used stimuli that varied along two dimensions. This resulted in an MDS space in which each stimulus was spatially represented along two dimensions with the distances between them representing the average dissimilarity obtained from the spatial arrangement task.

We ensured that the resulting MDS space was meaningful by using random stress and permutation tests (Mair, Borg & Rusch, 2016). In the random stress test, random dissimilarity data was constructed for 25 items and two-dimensional ordinal MDS was fitted to this random data. This process was repeated 10.000 times. Stress from our empirical MDS solution was compared to the average stress from the random MDS solutions, with the difference considered significant if observed stress was smaller than the lower 3 × *SD* boundary of the average random stress (Mair, Borg & Rusch, 2016). In the permutation test, our empirical dissimilarity data for the 25 stimuli was permuted to distort the underlying structure, and subsequently subjected to two-dimensional ordinal MDS. This process was repeated 10.000 times too, resulting in a null distribution of stress to which the stress value of our empirical MDS space was compared.

The MDS space is presented in Figure 2. The random stress test indicated that our data differed from random data, as the stress of our empirical MDS space (stress-1 = 0.09) was smaller than the lower 3 × *SD* boundary of the random stress norm (stress-1 = 0.30). The result of the permutation test was significant ($p < .001$), indicating that our MDS configuration can be discerned from configurations based on random permutations of the dissimilarity data.

As can be seen in Figure 2, the participants' spatial arrangements successfully captured the underlying length and color dimensions along which the stimuli varied, with coats varying in color from blue to purple along the x-axis, and varying in length from long to short along the y-axis. Because the MDS space looked as expected given the underlying dimensions of the images, it could safely be used for the external unfolding procedure.

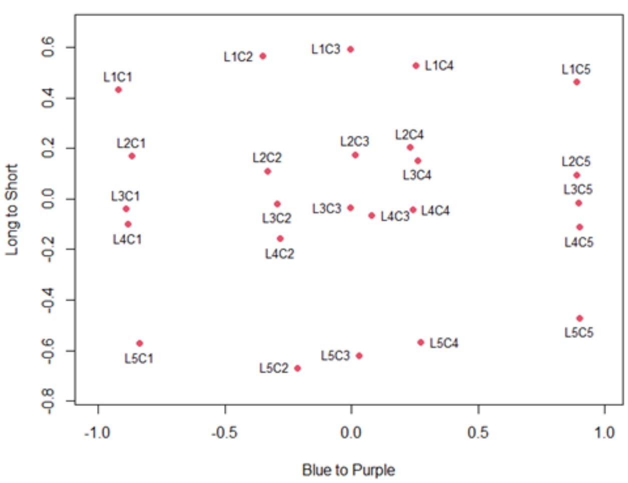

Figure 2: MDS space

External Unfolding

The reliability of the typicality data was estimated by applying the Spearman Brown formula to the split-half correlation, resulting in a reliability of .99 for each conjunction. As suggested by Barsalou (1987), however, the average interindividual correlations indicated a lot of variability between individuals, with average correlations of .62 (*SD* = .15) for *long blue coat*, and .57 (*SD* = .16) for *long purple coat*.

In order to represent the typicality data within the created MDS space, we performed two-dimensional ordinal external unfolding with the average and individual typicality data and the coordinates from our MDS space as input, using the VIPSCAL package (Van Deun, Groenen & Delbeke, 2005) in MATLAB R2023a (The Mathworks Inc., 2023). The ordinal level was again chosen so as to impose the fewest number of restrictions on the data. This procedure resulted in two unfolding solutions, one for *long blue coat*, and one for *long purple coat*, with participants represented as a vector and stimuli represented as in the MDS space.

General fit indices showed that the unfolding model fit the typicality data well, with $R = .80$, and $C1 = .91$ for *long blue coat*, and $R = .76$, and $C1 = .89$ for *long purple coat*.[3]

The unfolding solutions are presented in Figure 3. As can be seen, the unfolding procedure was successful in that most participants' vectors project into the expected quadrants. For example, for *long blue coat*, most vectors project into the lower left quadrant which is where the longer and more blue coats are located within the MDS space. These vector projections also indicate that performance in the typicality survey was as expected. For example, none of the participants rated short purple coats as highly typical of the conjunction *long blue coat*. Figure 3 also reveals a lot of variability between participants, with vectors distributed between horizontal and vertical projections. As can also be seen in Figure 3, the dotted black vectors representing the average typicality ratings are fairly representative of the average participant, in the sense that the vectors for the averages are situated in the middle of the spread of the vectors representing individual participants.

Discussion

This exploratory study aimed to use external unfolding to investigate whether evidence for concept dominance could be found for adjective-adjective-noun combinations, with a specific focus on individual differences between participants.

In line with the modest average interindividual correlations we found, the unfolding solutions revealed considerable variability in the participants' typicality judgements. This can clearly be seen from the variability in the directions of the vectors representing individual participants in Figure 3. Just as Barsalou (1987) suggested, the high reliability found for typicality ratings did not signal between-subject agreement, but rather the stability of the mean. Looking at our unfolding solutions, it becomes clear that this stability is due to the individual vectors clustering around the average vector (which is close to the diagonal) and becoming sparser as one moves farther away from the diagonal. The presence of these individual differences allows us to further investigate the differences between participants regarding concept dominance.

[3] *R* represents the average correlation to the typicality data. *C1* represents the average proportion of recovered preference orders.

Concept dominance refers to the finding that the relative contribution of constituent typicality to conjunction typicality is often very unequal (Hampton, 1988). So far, research into this effect has been conducted at the aggregate level (Hampton, 1988, 1997; Storms et al., 1993, 1996). Our results did not show a strong effect of concept dominance for the average typicality ratings. This can be seen in Figure 3, as the dotted black vectors representing the average typicality ratings project into the middle between the dimensions (i.e., the vectors roughly correspond to the diagonal). Slightly more weight appears to be given to the length dimension. The weight differences between the constituents were however not as pronounced as those reported in earlier research (Hampton, 1988, 1997; Storms et al., 1993, 1996). A possible reason for this could be that our stimuli, unlike those in earlier work where the stimuli were not under stringent experimental control, were selected in such a way that they varied along both underlying dimensions in equal amounts, which could have influenced participants to use both dimensions (roughly) equally. This raises the concern that the observation of concept dominance in earlier work might be an artifact of the stimulus selection procedure.

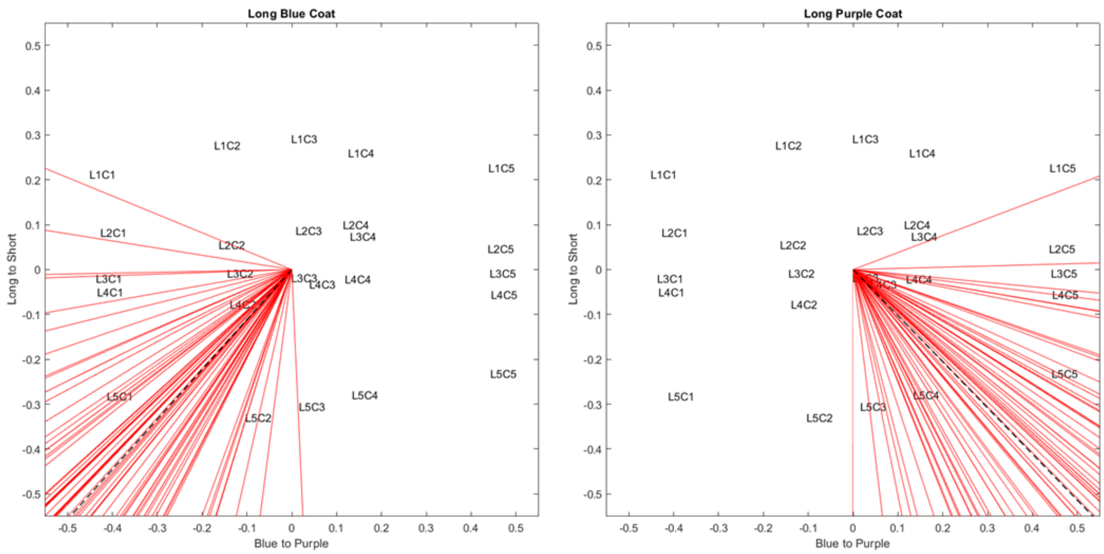

Figure 3: Unfolding solutions

In contrast to these findings at the aggregate level, our data did show evidence for concept dominance at the individual level, with vector projections revealing that some participants strongly favored one constituent dimension in each of the conjunctions. This can be seen in Figure 3, where some participants only seem to use the length dimension (as seen by their almost perfectly vertical vectors)[4] and other participants only seem to use the color dimension (as seen by their almost perfectly horizontal vectors) while rating images for typicality in the conjunction. Obviously, averaging across these "extreme" participants would result in an intermediate representation where it looks like participants made use of both dimensions equally (i.e., a null finding in terms of concept dominance). Alternatively, if there are more "extreme" participants favoring for example the length dimension than "extreme" participants favoring the color dimension, averaging will suggest a dominance effect of length *within* (rather than between) participants. However, neither of these outcomes would accurately reflect the way in which the typicality ratings were actually performed by these individuals, and this could lead to wrong conclusions about behavioral patterns in the population (Martin & Caramazza, 1980; Verheyen & Storms, 2007).

Of course, our sample did not merely consist of "extreme" participants. Most participants either showed no strong preference between dimensions, or favored the dimension that was also given slightly more weight in the average typicality ratings. However, there is still a lot of variability in which dimensions participants favored, and the degree to which they favored them. As a result, the average typicality ratings are not very representative for the behavioral patterns of many individuals within the sample. Even relatively small differences in participants' vector projections could represent meaningful differences in how they made typicality judgements. Through averaging, these more subtle differences between participants will also be obscured. For these reasons, our findings argue against an over-generalization of the notion of concept dominance to all individuals. Rather, our data indicate considerable variability in concept dominance between individuals and suggest that conceptual combination might be better explored at the individual level (taking into account both shared and

[4] The most extreme case of using only the length dimension was the participant with a visual impairment that affects the ability to distinguish between blue and purple (tritanomaly). However, examples of extreme concept dominance can also clearly be seen in other participants.

idiosyncratic sources of conceptual information; Barsalou, 1989), with pronounced individual differences in the extent to which constituent concepts are favored in this process.

A limitation of the present study is the assumption that the MDS space, obtained using averaged dissimilarity data, is shared by all participants. We cannot exclude the possibility that there were individual differences in participants' underlying similarity spaces. For example, some participants could give more weight to certain dimensions in judging item similarity. Because of this, it could be possible that the aggregate MDS space was not (entirely) appropriate to be used for external unfolding (Ashby, Maddox & Lee, 1994). One way to address this limitation in future research would be to use the INDSCAL algorithm (Carrol & Chang, 1970). INDSCAL is an MDS method that takes into account individual differences in the weight given to underlying dimensions. Using INDSCAL with the present methodology would allow the procedure to be repeated using individual MDS spaces.

Another limitation is that typicality ratings were collected using a 7-point Likert scale. It is possible that there were subtle differences between participants in their use of this scale. The present study did not control for this, therefore it is possible that some of the individual differences we found could be attributed to differences in scale usage.

Our finding that external unfolding is a useful technique for modeling individual differences in conceptual combination comes with exciting possibilities for theoretically motivated future research. The paper in which Hampton (1988) first reported concept dominance also contained findings on the noncommutativity effect, for instance. Noncommutativity refers to the order of the presentation of constituents having an effect on the relative weight with which they contribute to predicting typicality within the conjunction, with greater weight given to the constituent in the relative noun position. That is, constituent *B* has a larger weight than constituent *A* in the conjunction *A that is also B*. This noncommutativity effect too is well-established and replicated (Hampton, 1988, 1997; Storms et al., 1993, 1996). We were unable to investigate noncommutativity in the present study, as we always presented constituents in the natural order for English (length color noun). However, presenting participants with both orders would make it possible to investigate individual differences regarding the noncommutativity effect.[5]

Hampton (1997) also investigated conceptual combination in the context of negated conjunctions, for example by predicting typicality for *sports that are not games* from typicality for the constituents *sport* and *game*. These negated conjunctions also showed concept dominance. The present method can easily be adapted to investigate whether our findings can be generalized to negated conjunctions of adjectives such as *long coats that are not blue*.

Although a full investigation into intraindividual differences in typicality is beyond the scope of the present study, our data does show some behavioral patterns worth noting. Some participants, for instance, seemed to use only the size dimension for *long blue coat*, but relied almost entirely on the color dimension for *long purple coat*. Of course, these findings might be related to the different conjunctions. However, they could also be an indication that at least for some participants, the dimensions used to judge typicality are not stable over time. Verheyen, White & Egré (2019) found that the criteria participants used to make category membership decisions frequently differed between the first time they performed a categorization task, and the second time, when they performed the exact same task a month later. Other research with repeated measures has also revealed intraindividual differences for categorization (McCloskey & Glucksberg, 1978) as well as typicality (Barsalou, 1987, 1989). Of course, the timeframe in our study was much shorter, and participants never repeated the same exact task. However, these findings still raise the question of whether participants might for example use the length dimension while making typicality judgements for *long blue coat* on one day, and then mostly use the color dimension on another day. The present method can easily be used to investigate whether the criteria people use for typicality in adjective-adjective-noun combinations are subject to change over a longer timeframe by having them make the same typicality judgments on different occasions. In this manner it can address the question whether the individual differences we have observed in our study reflect stable interindividual differences that present repeatedly over occasions or whether they reflect the probabilistic (intraindividual) nature of the semantic retrieval process that underlies typicality ratings.

Conclusion

The present study did not find strong evidence for concept dominance (Hampton, 1988) in adjective-adjective-noun combinations at the aggregate level. However, evidence for concept dominance was found at the individual level, with marked individual differences in the extent of dominance and the dominant dimension. The study validates external unfolding for research on concept combination and comes with exciting new research possibilities for investigating noncommutativity, negated conjunctions, and intraindividual differences in concept representation.

Acknowledgments

SV conceptualized the study. WJL, SJ, and SBCJ collected and independently analyzed the data. The paper is based on the bachelor thesis by WJL. SJ, SBCJ, and SV provided critical revisions. All authors discussed the findings

[5] There is a caveat to this idea, as it has been found that English speakers prefer the conventional order of adjectives (*long blue coat*) over the alternative order (*blue long coat*) (Rosales & Scontras, 2019; Trainin & Shetreet, 2020). Noncommutativity is therefore probably better investigated in languages with less robust adjective preference orders, such as Hebrew, Spanish, and Arabic (Scontras, 2023; Trainin & Shetreet, 2023).

thoroughly, read, and approved the final version of the manuscript. They would like to thank Elisa Scerrati and Gabija Urbonaviciute for making the stimuli and James Hampton and Paul Egré for feedback and discussion on an earlier draft.

References

Ashby, F. G., Maddox, W. T., & Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science, 5*(3), 144–151.

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization*. Cambridge University Press.

Barsalou, L. W. (1989). Intraconcept similarity and its implications for interconcept similarity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. Cambridge University Press.

Borg, I., & Groenen, P. J. (2005). *Modern multidimensional scaling: Theory and applications.* Springer Science & Business Media.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika, 35*(3), 283–319.

de Leeuw, J., & Mair, P. (2009). Multidimensional Scaling Using Majorization: SMACOF in R. *Journal of Statistical Software, 31*(3), 1–30.

Goldstone, R. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers, 26*(4), 381–386.

Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition, 15*(1), 55–71.

Hampton, J. A. (1988). Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 12–32.

Hampton, J. A. (1996). Conjunctions of visually based categories: Overextension and compensation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(2), 378–396.

Hampton, J. A. (1997). Conceptual combination: Conjunction and negation of natural concepts. *Memory & Cognition, 25*(6), 888–909.

Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient, spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General, 142*(1), 256–281.

Mair, P., Borg, I., & Rusch, T. (2016). Goodness-of-fit assessment in multidimensional scaling and unfolding. *Multivariate Behavioral Research, 51*(6), 772–789.

Mair, P., Groenen, P. J. F., & de Leeuw, J. (2022). More on multidimensional scaling and unfolding in R: smacof Version 2. *Journal of Statistical Software, 102*(10), 1–47.

Martin, R. C., & Caramazza, A. (1980). Classification in well-defined and ill-defined categories: Evidence for common processing strategies. *Journal of Experimental Psychology: General, 109*(3), 320–353.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition, 6*(4), 462–472.

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from https://support.pstnet.com/.

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rosales, C. M., & Scontras, G. (2019). On the role of conjunction in adjective ordering preferences. *Proceedings of the Linguistic Society of America, 4*(1), 1–12

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

Scontras, G. (2023). Adjective ordering across languages. *Annual Review of Linguistics, 9*(1), 357-376.

Storms, G., De Boeck, P., Van Mechelen, I., & Geeraerts, D. (1993). Dominance and noncommutativity effects in concept conjunctions: Extensional or intensional basis? *Memory & Cognition, 21*(6), 752–762.

Storms, G., De Boeck, P., Van Mechelen, I., & Ruts, W. (1996). The dominance effect in concept conjunctions: Generality and interaction aspects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1266–1280.

Storms, G., Ruts, W., & Vandenbroucke, A. (1998). Dominance, overextensions and the conjunction effect in different syntactic phrasings of concept conjunctions. *European Journal of Cognitive Psychology, 10*(4), 337–372.

The MathWorks Inc. (2023). MATLAB version: 9.14.0 (R2023a), Natick, Massachusetts: The MathWorks Inc.

Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika, 17*, 401–419.

Trainin, N., & Shetreet, E. (2020). It's a dotted blue big star: On adjective ordering in a post-nominal language. *Language, Cognition and Neuroscience, 36*(3), 320–341.

Trainin, N., & Shetreet, E. (2023). Subjectivity predicts adjective ordering preferences in Hebrew, but lexical factors matter too. *Language, Cognition and Neuroscience, 38*(7), 1020-1026.

Tucker, L. R. (1960) Intra-individual and inter-individual multidimensionality. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and applications*. Wiley

Van Deun, K., Groenen, P. J. F. & Delbeke, L. (2005). VIPSCAL: A combined vector ideal point model for preference data. *Econometric Institute Report EI 2005-03*, Rotterdam: Erasmus University Rotterdam.

Verheyen, S., Egré, P., Hampton, J. A., Scerrati, E., & Urbonaviciute, G. (2022, September 20). Conjunctions of vague adjectives: within-participants design. Retrieved from osf.io/cge9j

Verheyen, S., Hampton, J., & Storms, G. (2010). A probabilistic threshold model: Analyzing semantic categorization data with the Rasch model. *Acta Psychologica*, *135*(2), 216-225.

Verheyen, S., & Storms, G. (2007). Modeling individual differences in learning hierarchically organised categories. *Psychologica Belgica, 47*(4), 219–234.

Verheyen, S., Voorspoels, W., Vanpaemel, W., & Storms, G. (2016). Caveats for the spatial arrangement method: Comment on Hout, Goldinger, and Ferguson (2013). *Journal of Experimental Psychology. General, 145*, 376-382.

Verheyen, S., White, A., & Egré, P. (2019). Revealing criterial vagueness in inconsistencies. *Open Mind: Discoveries in Cognitive Science, 3*, 41-51.

Verheyen, S., White, A., & Storms, G. (2022). A comparison of the Spatial Arrangement Method and the Total-Set Pairwise Rating Method for obtaining similarity data in the conceptual domain. *Multivariate Behavioral Research, 57*(2-3), 356-384.

Voorspoels, W., Vanpaemel, W., & Storms, G. (2011). A formal ideal-based account of typicality. *Psychonomic Bulletin & Review*, *18*(5), 1006–1014.