Silvey, C., Gentner, D., Richland, L., & Goldin-Meadow, S. (2017). Children's specific comparisons from 26 to 58 months predict performance in verbal and non-verbal analogy tests in 6th grade. To appear in G. Gunzelmann, A. Howes, T. Tenbrink & E. Davelaar (Eds.) *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 1072-1077). Austin, TX: Cognitive Science Society.

Children's spontaneous comparisons from 26 to 58 months predict performance in verbal and non-verbal analogy tests in 6th grade

Catriona Silvey (casilvey@uchicago.edu)
Department of Psychology, 5848 S. University Avenue
Chicago, IL 60637 USA

Dedre Gentner (gentner@northwestern.edu)
Department of Psychology, 2029 Sheridan Road
Evanston, IL 60208 USA

Lindsey Richland (lrichland@uchicago.edu)
Department of Comparative Human Development, 1126 E. 59th St.
Chicago, IL 60637 USA

Susan Goldin-Meadow (sgm@uchicago.edu)
Departments of Psychology and Comparative Human Development, 5848 S. University Avenue/1126 E. 59th St.
Chicago, IL 60637 USA

Abstract

Comparison supports the development of children's analogical reasoning. The evidence for this claim comes from laboratory studies. We describe spontaneous comparisons produced by 24 typically developing children from 26 to 58 months. Children tend to express similarity before expressing difference. They compare objects from the same category before objects from different categories, make global comparisons before specific comparisons, and specify perceptual features of similarity/difference before non-perceptual features. We then investigate how a theoretically interesting subset of children's comparisons – those expressing a specific feature of similarity or difference – relates to analogical reasoning as measured by verbal and non-verbal tests in 6th grade. The number of specific comparisons children produce before 58 months predicts their scores on both tests, controlling for vocabulary at 54 months. The results provide naturalistic support for experimental findings on comparison development, and demonstrate a strong relationship between children's early comparisons and their later analogical reasoning.

Keywords: comparison; similarity; language development; analogy

Introduction

Comparison – the process of jointly examining two objects or events and assessing their similarities and differences – is crucial in the development of children's word learning, categorization, and analogical reasoning skills (Namy & Gentner, 2002; Gentner & Namy, 2006; Gentner, Anggoro, & Klibanoff, 2011; Richland & Simms, 2015). Comparison is an effective learning tool because it promotes *structural alignment*: the mapping of two representations in a way that enables the recognition of relational commonalities and alignable differences. A large body of experimental work shows that inviting children to compare exemplars helps them to move beyond overall or global similarity to more specific kinds of similarity, including similarity based on relational commonalities, as in analogical reasoning (Loewenstein & Gentner, 2001; Christie & Gentner, 2014; Gentner et al., 2016). However, to get a full picture of the role of comparison in the development of children's analogical reasoning skills, it is important to relate this experimental work to children's spontaneous behavior in a naturalistic environment. Previous work has shown that children spontaneously produce comparative utterances from early in their language development: for example, children spontaneously generate metaphors from the age of around 2 (Winner, 1979) and are able to explain them in terms of similarity (Billow, 1981). However, the nature of the comparisons children produce is not static over time, but follows a developmental trajectory. Özçalışkan, Goldin-Meadow, Gentner, and Mylander (2009) found that while children's earliest comparisons tended to be between objects that were similar to each other in many features, the acquisition of the word 'like' was associated with an increase in the number of comparisons between objects that only shared a single feature. These specific comparisons are argued to be a more sophisticated stage in the development of children's understanding of similarity than are global comparisons (Smith, 1989; Gentner & Rattermann, 1991). As such, the prevalence of specific comparisons in children's early speech could potentially be an index of their later analogical reasoning skill.

The current work has two aims: 1) a descriptive aim, to characterize common patterns in the development of children's spontaneous comparisons produced in naturalistic contexts in the home; 2) an inferential aim, to test the hypothesis that variation in children's production of specific, single-feature comparisons predicts variation in their scores on tests of analogical ability given much later, in 6th grade.

Methods

Participants

24 children and their primary caregivers were drawn from a larger sample of 64 families who participated in a longitudi-

nal study of language development (the same sample drawn on by Özçalışkan et al., 2009). Families were recruited via direct mailings to targeted zip codes and an advertisement in a free monthly parenting magazine. Parents who responded were interviewed regarding background characteristics, and the final sample was selected to be representative of the greater Chicago area in terms of race, ethnicity and income. The sub-sample of 24 families in the current study was selected randomly, within the constraints of preserving the demographic spread of the original sample. Of the 24 children, 11 were male and 13 female; 18 were white (of whom 3 were Hispanic), 3 were Black or African-American, and 3 were of two or more races. The distribution of socio-economic status across the 24 families was similar to that of the original sample, ranging from families with an income of under $15,000 where the primary caregiver had some high school education, to families with an income of over $100,000 where the primary caregiver had an advanced degree.

Procedure

Parents and children were visited in their homes and videotaped engaging in their normal daily activities for 90 minutes. Home visits began when the children were 14 months old and continued at 4-month intervals, ending when the children were 58 months old (12 sessions in total).[1] All child speech, and all parent speech directed to the child, was transcribed. Transcription reliability was established by having a second individual transcribe 20% of each transcriber's tapes. Reliability was at or above 95%.

Coding

Comparisons were coded from the transcripts of child speech during the 12 sessions. The criterion for a comparison was that the child expressed a similarity or difference between an identifiable source and target. Sources and targets could be objects or events. In cases where the source and target of the comparison were unclear from the transcript alone, the original video was consulted. For each identified comparison, we coded the following:

Word. The word that made the utterance a comparison; e.g. 'I'm a funny one like you' would be coded as 'like'.

Word category. Comparative words were classified into six categories: like (the words 'like' and 'alike'), same/different (the words 'same' and 'different'), comparative/superlative (any comparative or superlative adjective, e.g., 'bigger', 'best'), too (used either in contexts like 'too big' or contexts like 'I'm dancing too'), match (e.g., 'these match each other'), and other.

Object or event. Comparisons were coded for whether the Source and Target were objects (e.g., 'this [rug] look like a skirt') or events (e.g., 'I win too').

Expressing similarity or difference. Comparisons were coded for whether they expressed similarity (e.g. 'go like a elephant') or difference (e.g. 'I'm bigger than everybody!').

Global or specific comparison. Comparisons were coded for whether they expressed global similarity/difference (e.g., for Objects, 'I have toys just like yours'; for Events, 'they both win'), or specific similarity/difference (e.g., for Objects, 'red like the ladybug'; for Events, 'I go a lot faster than when I was three'). Comparisons could be specific even if the objects compared were overall similar, e.g., 'this [tree] is the tallest [tree]'. We expect global comparisons to appear earlier than specific comparisons (Smith, 1989; Gentner & Rattermann, 1991).

Feature specified. Where a feature of similarity or difference was specified, this feature was coded. Features were classified into 6 categories: Spatial (e.g., size, shape, distance, speed), Sensory (e.g., color, weight, taste, smell), Evaluative (e.g., goodness, prettiness, badness), Emotion (e.g., being tired, mad, scared), Preference (e.g., liking one thing better than another thing),[2] and Other. Features were also classified as Perceptual (based on a readily perceptible attribute, e.g. color, size) or Non-Perceptual (based on a more abstract, not directly perceptible feature, e.g., preference, goodness).

Within or between-category comparison. Comparisons were coded for whether the objects compared were from the same or different superordinate categories. Superordinate categories were taken from Özçalışkan et al. (2009), with three additions to accommodate new data (in italics): people, animals, body parts, vehicles, clothing, furniture, appliances, kitchen utensils, tools, musical instruments, food, plants, activity toys, places, *decorations/crafts*, *words/letters*, and *shapes*.

In the case of events, the objects of interest were those with corresponding roles in the two events. For example, if the parent said she was going to use some yellow paint, and the child said 'think I'll do yellow too', the objects in corresponding roles (parent/child, and yellow paint/yellow paint) are in the same superordinate categories (people and decorations/crafts, respectively). This would therefore be coded as a within-category comparison. If the child said 'I'm going to act like a bee', the objects in corresponding roles (child and bee) are in different superordinate categories (people and animals); this would therefore be coded as a between-category comparison. If children initially rely on overall similarity, then within-category comparisons should emerge earlier than between-category comparisons.

A total of 532 comparisons were codable under these guidelines.

Later outcomes

The same children were followed longitudinally as part of an ongoing language development project. When the chil-

[1] Since no comparisons were produced before session 4 (26 months), graphs & analyses focus on sessions 5-12 (26-58 months).

[2] Utterances using the word 'favorite' were not coded, since it was not clear that children understood its meaning as comparative.

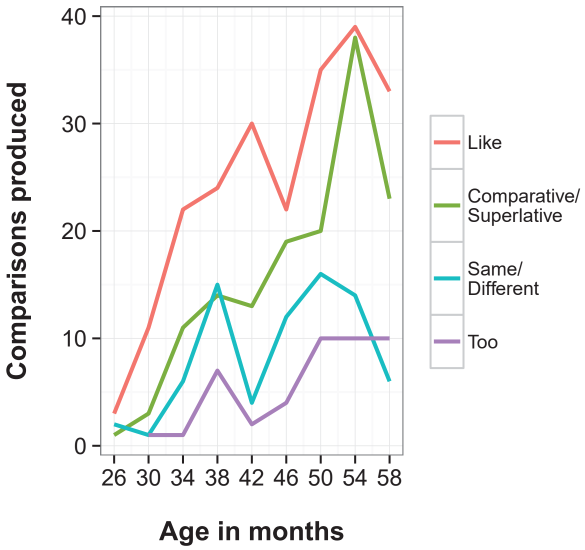

Figure 1: Frequencies of word categories across sessions.

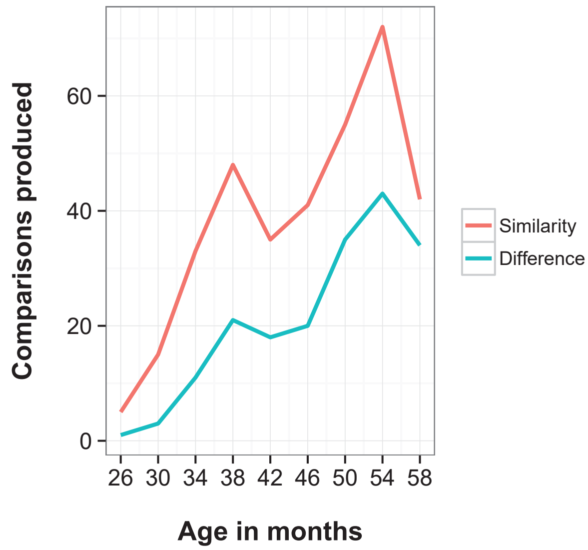

Figure 2: Frequencies of comparisons expressing similarity and difference across sessions.

dren were in 6th grade (aged around 13 years), we administered two tests of analogical reasoning: the Verbal Analogies subtest of the Woodcock-Johnson Tests of Cognitive Abilities (Woodcock, McGrew, & Mather, 2001), and a nonverbal test, Raven's Progressive Matrices (Raven, 1938). The Woodcock-Johnson Verbal Analogies is an orally administered test that consists of sets of paired items. The participant has to fill in the missing item by abstracting the relation that holds between the first pair. For example, the participant is given the prompt 'mother is to father, as sister is to...', and expected to fill in the missing term 'brother'. Raven's Progressive Matrices consists of a series of geometric analogy problems. The participant is presented with a matrix that has one entry missing and must select the correct entry from an array of 6-8 choices. These two measures were taken as outcomes in our analyses.

Results

Onset and prevalence of comparisons

Children varied in the age at which they produced their first comparison. For the purpose of this analysis, age of onset was defined as children's age during the session where they produced at least one comparison and also produced at least one comparison during the immediately following session. Under this criterion, the earliest onset was at 26 months, and the latest was at 50 months. The average age of onset was 36 months, with a standard deviation of 6 months. Comparisons were relatively infrequent: they ranged from 0% to 2.2% of a child's utterances in a given session. However, the fact that we reliably find comparisons even in short 90-minute sessions suggest they are a robust feature of children's early talk.

Comparison words

The most commonly used comparison word was 'like', followed by 'too', 'bigger', 'same', and 'better'. Together, these words accounted for 73% of the comparisons the children expressed. Table 1 shows counts and percentages for the word categories detailed in the Methods.

Table 1: Word categories.

Word category	Number of uses	Percent
like	219	41%
comparative/superlative	142	27%
too	76	14%
same/different	45	8%
other	34	6%
match	16	3%

Figure 1 shows the frequencies of the 4 most prevalent word categories over sessions. 'Like' is the first word category to reliably emerge. While 'like' and comparatives/superlatives are overall more frequent, all word categories generally show an increase in use across sessions.

Expressing similarity and difference

Figure 2 shows the trend over sessions for expressing similarity versus difference. Similarity comparisons were more numerous overall (346 to 186). The general trend was for similarity comparisons to emerge earlier than difference comparisons, and to remain more numerous until the final session. On a by-individual level, 20 out of 24 children produced a similarity comparison before they produced a difference comparison; 1 produced a difference comparison before produc-

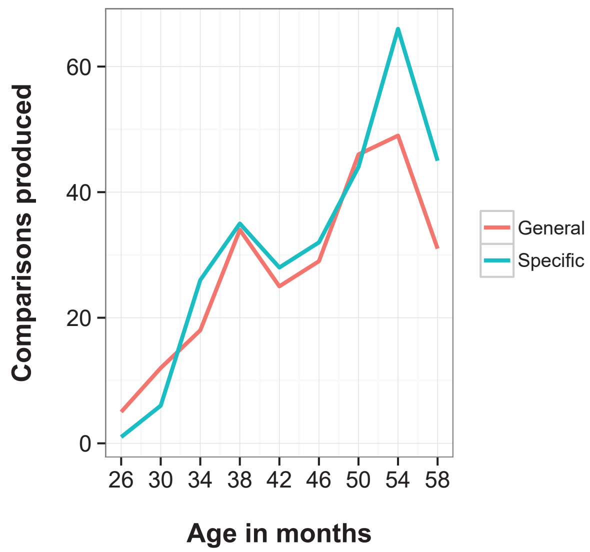

Figure 3: Frequencies of global and specific comparisons across sessions.

ing a similarity comparison; and 3 produced examples of both simultaneously. This trend for similarities to precede differences was significant, $\chi^2 = 27.25, p < .001$.

Objects and events

While object comparisons were more numerous in general (358 compared to 174 event comparisons), the overall trend was for object and event comparisons to emerge at around the same time. 11 out of 24 children produced an object comparison before they produced an event comparison; 8 produced an event comparison before they produced an object comparison; and 5 produced examples of both simultaneously. The trend in ordering was not significant, $\chi^2 = 2.25, p = .32$. Thus it appears that from comparison onset, children are capable of expressing comparisons between events as well as comparisons between objects.

Global and specific comparisons

The numbers of global and specific comparisons were broadly equivalent: 249 global to 283 specific. Figure 3 shows the trend over sessions. Global comparisons appear to be more numerous than specific comparisons in the first two sessions; in subsequent sessions they are at equivalent levels, until the final two sessions when specific comparisons are higher. By individuals, as predicted, global comparisons tended to precede specific comparisons: 14 of 24 children produced a global comparison before they produced a specific comparison, while 5 produced a specific comparison before they produced a global comparison, and 5 produced both in the same session. While not as strong as the tendency for similarity to precede difference, this trend in ordering was significant, $\chi^2 = 6.75, p = .034$.

Features specified

The most frequently specified features were spatial or sensory; together, these accounted for 70% of the specific comparisons the children expressed. Table 2 shows overall counts and percentages.

Table 2: Feature categories.

Feature category	Number of uses	Percent
Spatial	136	48%
Sensory	62	22%
Evaluative	49	17%
Other	30	11%
Emotion	4	1%
Preference	3	1%

More perceptual features (202) were specified than non-perceptual features (80). The overall trend was for perceptual features to be specified earlier: by individual, 16 children specified perceptual features before they specified non-perceptual features, 4 specified non-perceptual features before they specified perceptual features, and 4 did both in one session. The trend for perceptual features to be specified first was significant, $\chi^2 = 12, p = .002$.

Within- and between-category comparisons

Comparisons between objects in the same superordinate category (or between events involving objects in the same superordinate categories) were more numerous than comparisons between different superordinate categories (421 compared to 133). As predicted, comparisons between objects in the same category generally tended to precede comparisons between objects in different categories. 14 of 24 children produced a within-category comparison before a between-category comparison. 5 produced a between-category comparison first, and 5 children did both in one session. This trend in ordering was significantly different from chance, $\chi^2 = 6.75, p = .034$.

Comparison type interactions

We also examined interactions between comparison types.

Firstly, we asked whether the children's comparisons expressing similarity were more likely to specify a feature than their comparisons expressing difference, or vice versa. 118 (34%) of similarity comparisons specified a feature of similarity, while 165 (89%) of difference comparisons specified a feature of difference. Given their marginal totals, similarity comparisons were less likely than expected to specify features, and difference comparisons were more likely than expected to specify features. This difference was significant, $\chi^2 = 145.38, p < .001$.

We then asked whether comparisons involving objects in the same superordinate category were more likely to express similarity or difference, as opposed to comparisons involving objects in different superordinate categories. Comparisons of within-category objects, or events involving within-category

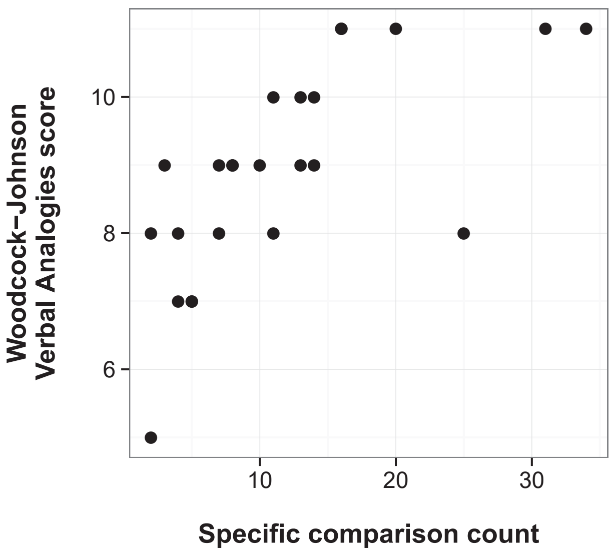

Figure 4: Scatterplot showing number of specific comparisons produced from 26-58 months (x axis) and score in Verbal Analogies test in 6th grade (y axis).

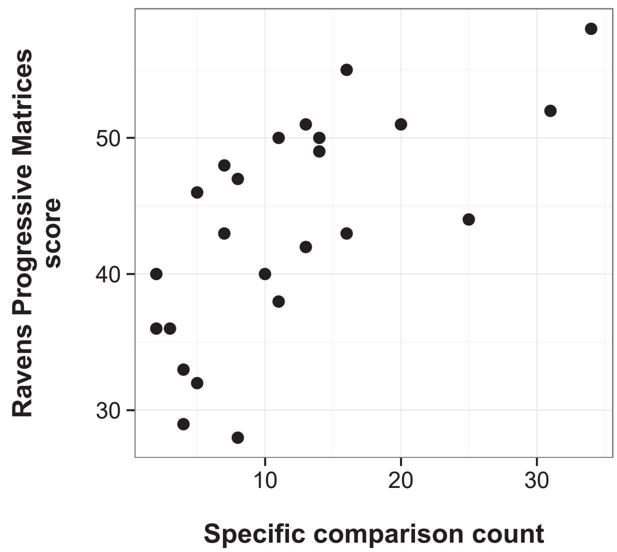

Figure 5: Scatterplot showing number of specific comparisons produced from 26-58 months (x axis) and score in Raven's Progressive Matrices test in 6th grade (y axis).

objects, were broadly as likely to express similarity as difference: 240 (60%) of these expressed similarity. On the other hand, comparisons of between-category objects, or events involving between-category objects, were more likely to express similarity (105, or 80%) than difference. This trend was significant, $\chi^2 = 15.32, p < .001$.

Relation to later outcomes

We then tested the hypothesis motivated in the Introduction, that the number of specific comparisons (expressing a single feature of similarity or difference) that children made during the 12 observational sessions would predict their performance on tests of analogical reasoning in 6th grade.

Our outcome measures were the two analogy tests described in the Methods: the Woodcock-Johnson Verbal Analogies test, and Raven's Progressive Matrices. Both a verbal and a non-verbal test were administered in order to address the potential confound of language skill, which could influence both children's comparison production and their verbal analogy test scores. To further account for language proficiency, we controlled for the child's score on the Peabody Picture Vocabulary Test (PPVT-III; Dunn & Dunn, 1997) at 54 months (the penultimate session of the 12 during which comparisons were collected).

Figures 4 and 5 show scatterplots of the relationship between the number of specific comparisons the children produced during the pre-school observation sessions and their 6th grade scores on the Verbal Analogies and Raven's Progressive Matrices tests, respectively.

Table 3 shows the results of the statistical model predicting Verbal Analogies score from specific comparison count and PPVT at 54 months. Specific comparisons remained a significant predictor after controlling for PPVT, although PPVT had a larger effect. The adjusted R^2 for the model was .64, indicating that these two variables together explain around two-thirds of the variance in Verbal Analogies score.

Table 3: Verbal Analogies model

Predictor	Standardized β	*t*	*p*
# specific comparisons	0.37	2.44	.024
PPVT at 54 months	0.55	3.60	.002

Table 4: Raven's Progressive Matrices model

Predictor	Standardized β	*t*	*p*
# specific comparisons	0.67	4.27	$< .001$

Table 4 shows the results of the model predicting Raven's Progressive Matrices score from specific comparison count. In this case, a likelihood ratio test showed that adding PPVT did not improve the model, $F(1) = 1.05, p = .318$. The adjusted R^2 for the model was .43, indicating that specific comparison count alone explains around 40% of the variance in Raven's Progressive Matrices scores.

Discussion

Children's earliest comparisons tend to express global similarity between objects or events within the same superordinate category. Later in development, children begin to express difference, to specify features of comparison, and to compare objects and events from different superordinate categories. Turning to the content of these comparisons, children are particularly motivated to comment first on similarities and differences in perceptual features such as size, color, and speed, and later on evaluative features such as goodness, prettiness, and their opposites.

While children are more likely to express global similarity than specific similarity, most difference comparisons are specific rather than global. This finding suggests that children are less motivated to comment on overall dissimilarity than on overall similarity: differences are only interesting insofar as they are specific. We also find that comparisons involving objects in different superordinate categories tend to disproportionately express similarity, rather than difference, despite these objects being a priori less similar to each other. This seemingly counter-intuitive result backs up existing theory: more similar objects are more likely to have salient, alignable differences than objects which are dissimilar (Markman & Gentner, 1993; Gelman, Raman, & Gentner, 2009).

The relationship we find between children's early comparisons and their later analogical reasoning skill can potentially be interpreted in a number of ways. One possibility is that children who make more specific comparisons gain more practice in identifying dimensions of similarity or difference: thus, making these comparisons directly helps build their analogical skills in ways that persist through later development. Another possibility is that both our predictor variable (the prevalence of specific comparisons in the pre-school years) and our outcome variable (performance on verbal and nonverbal analogy tests in 6th grade) can be traced back to an underlying variable such as intelligence. The current work cannot tease these explanations apart. However, in future work, we aim to code the comparisons parents produce during the sessions before their children start producing comparisons themselves. It will then be possible to use causal modeling to investigate the extent to which parent comparison input predicts child comparison production, controlling for parent IQ. If parent comparison input influences child production of comparisons beyond a heritable IQ effect, this outcome could potentially open the door for interventions aimed at boosting children's comparison production in the home by providing them with particularly helpful kinds of input.

Acknowledgments

Thanks to Anjali Murthy for help with coding scheme refinement. This research was funded by NIH P01 HD040605 to Goldin-Meadow.

References

Billow, R. M. (1981). Observing spontaneous metaphor in children. *Journal of Experimental Child Psychology*, *31*, 430–445.

Christie, S., & Gentner, D. (2014). Language helps children succeed on a classic analogy task. *Cognitive Science*, *38*, 383–397.

Dunn, L. M., & Dunn, L. M. (1997). *PPVT-III: Peabody Picture Vocabulary Test*. Circle Pines, MN: American Guidance Service.

Gelman, S. A., Raman, L., & Gentner, D. (2009). Effects of language and similarity on comparison processing. *Language Learning and Development*, *5*(3), 147–171.

Gentner, D., Anggoro, F. K., & Klibanoff, R. S. (2011). Structure mapping and relational language support children's learning of relational categories. *Child Development*, *82*(4), 1173–88.

Gentner, D., Levine, S. C., Ping, R., Isaia, A., Dhillon, S., Bradley, C., & Honke, G. (2016). Rapid learning in a children's museum via analogical comparison. *Cognitive Science*, *40*(1), 224–240.

Gentner, D., & Namy, L. L. (2006). Analogical processes in learning. *Current Directions in Psychological Science*, *15*(6), 335 – 361.

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Brynes (Eds.), *Perspectives on thought and language: Interrelations in development* (pp. 225–277). Cambridge: Cambridge University Press.

Loewenstein, J., & Gentner, D. (2001). Spatial mapping in preschoolers: Close comparisons facilitate far mappings. *Journal of Cognition and Development*, *2*(2), 189–219.

Markman, A. B., & Gentner, D. (1993). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language*, *32*(4), 517–535.

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General*, *131*(1), 5–15.

Özçalışkan, Ş., Goldin-Meadow, S., Gentner, D., & Mylander, C. (2009). Does language about similarity play a role in fostering similarity comparison in children? *Cognition*, *112*(2), 217–228.

Raven, J. C. (1938). Progressive matrices: A perceptual test of intelligence. *London: HK Lewis*.

Richland, L. E., & Simms, N. (2015). Analogy, higher order thinking, and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(2), 177–192.

Smith, L. B. (1989). From global similarities to kinds of similarities: The construction of dimensions in development. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 225–277). New York: Cambridge University Press.

Winner, E. (1979). New names for old things: The emergence of metaphoric language. *Journal of Child Language*, *6*(03), 469–491.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of cognitive abilities*. Itasca, IL: Riverside Publishing.