UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Structural Alignment Across Visual and Linguistic Modalities: A Developmental Refinement Perspective

Permalink
https://escholarship.org/uc/item/5cv4t5mf

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 47(0)

Author
Li, Luan

Publication Date
2025

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Structural Alignment Across Visual and Linguistic Modalities: A Developmental Refinement Perspective

Luan Li (liluan@sjtu.edu.cn)
Shanghai Jiao Tong University, Shanghai, China

Abstract

This study investigates the structural alignment of conceptual representations accessed from visual and linguistic modalities in 8-9-year-old children and adults. Using a Spatial Arrangement (SpAM) task, participants organized familiar items from two categories – household items (HHI) and vegetables & fruits (VF) – presented separately as images and written words on a two-dimensional grid. Representational dissimilarity matrices were computed based on item distances within each modality and analyzed for structural alignment across modalities. Results showed significantly lower cross-modal alignment in children compared to adults, suggesting ongoing developmental changes in the structure of conceptual representations. Additionally, cross-modal alignment was higher for VF than for HHI categories in both age groups, indicating category-specific variations in conceptual organization and its refinement. These findings provide insights into the gradual refinement of the structure of the lexical-conceptual system, extending beyond item-level lexical learning.

Keywords: cross-modal alignment; conceptual development; category; representational similarity analysis

Introduction

A central challenge in cognitive science is understanding how humans develop a unified conceptual system that aligns across input modalities, such as vision and language. From early childhood, this cross-modal alignment scaffolds word learning, building on foundational work showing that infants and toddlers integrate sensory inputs to form coherent concept knowledge (Smith & Yu, 2008). At its most basic level, conceptual word learning begins with mapping a linguistic label to its referent in the physical world. However, one-to-one label-referent mappings alone cannot fully explain phenomena such as contextual variability of word meaning (Lupyan & Lewis, 2019) or the role of cultural and individual experiences in shaping conceptual system (Wojcik, Zettersten, & Benitez, 2022). While early acquisition relies on item-level mappings (Woodward, Markman, & Fitzsimmons, 1994), developmental research increasingly highlights how lexical representations undergo prolonged structural refinement beyond early childhood (Li & Zhang, 2025; Vales, States, & Fisher, 2020; Vales, Stevens, & Fisher, 2020).

An alternative perspective to conceptual growth emphasizes the role of cross-modal structural alignment (Aho, Roads, & Love, 2023; Luo, Zhang, Xiao, & Brendan, 2024; Roads & Love, 2020). This framework suggests that conceptual learning involves not only mapping individual labels to referents but also acquiring higher-order structural similarities across systems, such as language and vision. For example, if the words "dog" and "cat" frequently co-occur in language, their corresponding concepts are also likely to co-occur in the visual domain. In this view, learning the relationships between items in one system (e.g., language) develops, at least partly, in tandem with the growth of these relationships in another system (e.g., vision), enabling global cross-modal alignment. Rather than first establishing item-level mappings and then building relationships between items, this proposal emphasizes the simultaneous development of item- and structural-level knowledge across systems, contributing to a unified conceptual network that encodes multimodal information. This perspective is particularly appealing because it recognizes that lexical knowledge extends beyond label-referent mappings and includes relationships between other words and structural information embedded in multimodal contexts.

Growing evidence supports the idea of system alignment in conceptual acquisition. For example, research has shown that children's early concepts show structural alignment across language (spoken and written) and visual systems (Roads & Love, 2020). Further, children's early concepts are densely connected, facilitating the learning of novel concepts through system alignment (Aho et al., 2023). Similarly, a study with adults has found that learning mappings between perceptual and spatial features for novel items is easier when the two systems share consistent structural mappings, compared to when they are misaligned (Aho, Roads, & Love, 2022). Broadly, map-like structural representations between concepts are conducive to flexible behaviors (Behrens et al., 2018) including lexical access (Borovsky, Ellis, Evans, & Elman, 2016) and creative thoughts (Kenett & Faust, 2019).

Building on these findings, we propose that the acquisition of novel concepts and the refinement of conceptual relationships evolve hand-in-hand during development. As new items are added into the conceptual system, its structure undergoes continuous refinement to accommodate both the new items and their relationships with existing ones (Hills & Siew, 2018; Sizemore, Karuza, Giusti, & Bassett, 2018). In this view, the global and local structures of cognitive systems are in constant flux, as is the extent to which they align with other systems. However, while recent research has highlighted how early concept learning is supported by system alignment, the development of this alignment itself remains poorly understood.

The current study examines the refinement of cross-modal structural alignment beyond early childhood, focusing on the structural similarity between concepts accessed through

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

language and visual systems that are readily acquired in school-age children and adults. We hypothesized that global structural changes within each system during development will influence local structures. Specifically, we expected that children would show less cross-modal alignment compared to adults, even for highly familiar concepts, as their language and vision systems' global structural representations continues to evolve.

Moreover, different sections of a conceptual system may develop at varying rates. Certain concepts and their connections are acquired earlier than others. For example, children learn vision verbs before other sensory verbs (San Roque, Norcliffe, & Majid, 2024). Concrete noun categories, which we examine in this research, also develop at different paces: basic categories (e.g., chair) tend to emerge first, followed by superordinate categories (e.g., furniture), and finally subordinate categories (e.g., rocking chair, Mervis & Crisafi, 1982). Young children also show an early preference for animate over inanimate concepts (Opfer & Gelman, 2010) and acquire food categories earlier than tool categories (Fenson et al., 1994).

This varying developmental trajectory may stem from differences in how categories weight different types of information in their representations. While all concrete noun categories primarily depend on visual information, natural kinds concepts rely more heavily on perceptual features whereas artifacts depend more on actions and functional features (Gainotti, Spinelli, Scaricamazza, & Marra, 2013). Additionally, children may have differential familiarity with these categories. For example, tools may be less salient and less frequently encountered than food, resulting in weaker representation of their functional features. Consequently, the role of visual modality may be less pronounced for tools than food, potentially leading to lower cross-modal alignment for tools despite comparable verbal familiarity.

We therefore hypothesized that the degree of cross-system alignment may vary across subsections of the conceptual systems. To test this, we probed children's and adults' structural representations of two semantic categories: Household Items, Vegetables & Fruits, which are typically acquired as early as toddlerhood (Fenson et al., 1994). Despite the maturity of their internal categorical structures, we predicted lower cross-modal alignment for Household Items than for Vegetables & Fruits.

Methods

Design

A Spatial Arrangement (SpAM) task was used to investigate the structural representations of conceptual relationships in children and adults. In the SpAM task, participants arrange multiple items in a two-dimensional space, with the distances between items reflecting their psychological representational distances. This task has been validated in prior research with children (e.g., León-Villagrá, Ehrlich, Lucas, & Buchsbaum, 2022).

Participants completed the SpAM task for the same sets of concepts presented in two modalities: as visual objects (i.e., visual modality) and as written names (i.e., language modality). Data from the visual modality have also been analyzed and reported for different research questions in Li & Zhang (2025).

To quantify the structural organization of the concepts, a Representational Dissimilarity Matrix (RDM) was computed for each modality. Cross-modality alignment scores were then calculated for each participant by determining the Spearman rank correlation between the RDMs of the same set of concepts in the visual and language modalities.

Participants

The study included 50 children in 2nd and 3rd grades (27 females, age M = 8.85, SD = 0.85) recruited from an elementary school, and 40 adults (25 females, age M = 20.4, SD = 1.50) recruited from a university. Data from one child was excluded due to incomplete task performance in the language modality. None of the children were reported by their teachers to have a history of language, hearing, or developmental disorders, and none of the adults reported such history. All participants provided informed consent before participation.

Stimuli

The stimuli consisted of 24 exemplar items from two categories: Household Items (HHIs) and Vegetables & Fruits (VFs). The HHIs included 6 kitchenware and 6 tools, and the VFs comprised 6 vegetables and 6 fruits. Visual depictions of the items and their names were sourced from the Chinese Image Set (CIS), a standardized psycholinguistic database (Ni, Liu, Yu, & Fu, 2019).

For each stimulus, two additional variables were calculated: 1) objective age of acquisition (AoA) of the item names, which was defined as the earliest age at which at least 50% of children produced the word, based on data from the CDI (Communicative Development Inventory) using the *wordbankr* R package (Braginsky, 2024). AoA values ranged from 16 to 30 months; 2) standardized frequency of item names in child-directed speech was calculated using data from the CHILDES corpus retrieved using the *childesr* package (Braginsky, Sanchez, & Yurovsky, 2019), which contains 1,424,459 word tokens.

Descriptive statistics for the stimuli are summarized in Table 1. The items were matched across categories (HHIs and VFs) on key psycholinguistic variables, including CDI and adult-reported AoAs, word frequency, familiarity, image visual complexity and adults' image naming response latencies. However, two-sample t-tests showed significant difference in several variables: image typicality, image agreement, category agreement and name agreement, with VF items scoring higher on all dimensions (ps < .05). These differences suggest that VF items had greater internal coherence in mental representations and more consistent object-label mappings compared to HHIs.

Table 1: Means (standard deviations in parentheses) of the psycholinguistic variables of HHI and VF stimuli.

	HHI	VF
Control variables		
Word AoA: CDI	1.71 (0.12)	1.69 (0.12)
Word AoA: adult-report	3.21 (0.57)	3.02 (0.34)
Word Frequency	1.66 (0.41)	1.62 (0.47)
Image Familiarity	3.84 (0.74)	4.15 (0.29)
Image visual complexity	1.90 (0.46)	2.20 (0.66)
Naming latency	1089.77 (253.82)	967.47 (122.41)
Other variables from CIS		
Image typicality	3.97 (0.45)	4.42 (0.58)
Image agreement	3.58 (0.59)	4.43 (0.41)
Image category agreement	0.69 (0.20)	0.98 (0.04)
Name agreement	0.79 (0.15)	0.95 (0.06)

Note. AoA: CDI *was calculated as the earliest age at which at least 50% of the children could produce the word;* AoA: adult-report *was from the CIS norm (Ni et al., 2019) where adults rated on "the age at which you learned the name of the item depicted in the photo"; Frequency was computed on 1,424,459 word tokens of child-directed speech from CHILDES using log10 (counts per million); The rest of the variables were from CIS, where adults rated on a 5-point scale on. Familiarity: "the degree to which you come in contact with or think about the concept"; Image visual complexity: "the amount of visual texture or intricacy of lines of the item depicted in the picture"; Image typicality: "the degree to which you think the item can represent the category it belongs to"; Image agreement: "the degree to which your mental image of the item matches the one depicted in the photo"; Category agreement: "the degree to which the object fits one of 21 specific categories"; Name agreement is the percentage of participants who provided the object name in a written picture naming task.*

Procedure

Participants were tested individually in a quiet room. Each participant was presented with a board containing an 8 × 8 grid of 4cm × 4cm squares. Twelve 4cm × 4cm cards, each depicting either an object or the written name of an item from one category, were placed beside the board.

Participants first examined each card individually and indicated any objects or names that they did not recognize. Unrecognized items were removed from the task.

Following the familiarity phase, the experimenter randomly placed one card on the grid and provided the following instruction: "Please place the cards on the board like I did. You can put them anywhere, but remember to put cards closer if you think they should go together."

Participants were allowed to rearrange the cards as many times as they wished and were given unlimited time to complete the task. After completing the arrangement of one category, participants proceeded to the other category in a randomized order. The order of the vision and language modalities was counterbalanced across participants.

Once participants confirmed their arrangements, the experimenter photographed the board for later analysis. The photos were coded by recording the x and y coordinates of each card on the 8 × 8 grid. Using these coordinates, Euclidean distances between each pair of items within a category and modality were calculated and normalized to a 0-1 range for each participant.

Results

Task reliability

The reliability of the SpAM task was assessed using split-half correlations. For each combination of age group (child, adult), modality (vision, language) and category (HHI, VF), participants were randomly divided into two halves. Distances between item pairs were averaged within each half, and Pearson correlations between the two halves were calculated and corrected using the Spearman-Brown formula. This process was repeated 100 times, resulting in a distribution of correlation coefficients for each age group, modality and category. Reliability was generally high in all conditions (Table 2), with adults consistently showing higher reliability than children. In the child group, the visual modality demonstrated slightly higher reliability than the language modality.

Table 2: Means (standard deviations in parentheses) of 100 repetitions of split-half correlations of the SpAM task by group, modality and category.

		Child	Adult
Vision	HHI	.90 (.02)	.94 (.02)
	VF	.89 (.03)	.93 (.02)
Language	HHI	.74 (.05)	.94 (.02)
	VF	.86 (.03)	.95 (.01)

Cross-modal structural alignment

Participant-level Analyses Cross-modal alignment was calculated at the participant level as the Spearman rank correlation between RDMs in the visual and language modalities. Individual structural alignment scores are presented in Figure 1.

To examine the effects of group (children vs. adults) and category (HHI vs. VF) on cross-modal alignment, a linear mixed effects (LME) model was fitted. The model included group and category fixed effects and a random intercept for subject. Results showed a significant effect of group (β = 0.28, SE = 0.05, t = 5.65, p < .001), indicating that adults had higher alignment scores than children. However, there was no significant main effect of category or interaction between group and category.

Pairwise comparisons between children and adults were conducted for HHIs and VFs, with Bonferroni correction applied for multiple comparisons. As predicted, children had significantly lower alignment scores than adults for both HHIs (child M = .14, SD = .19; adult M = .41, SD = .27; t =

-5.65, $p < .001$) and VFs (child $M = .17$, $SD = .28$; adult $M = .47$, $SD = .34$; $t = -5.65$, $p < .001$).

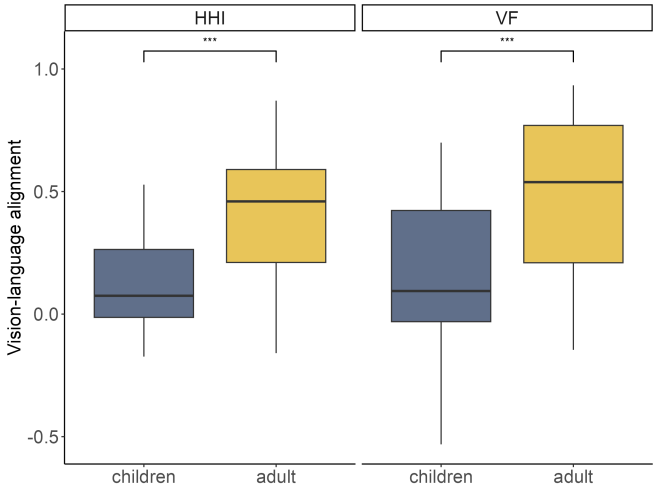

Figure 1: Boxplots of individual structural alignment scores across vision-language modalities in children and adults for household items and vegetables & fruits.

Group-level Analyses Given the variability in children' individual alignment scores, further analyses were conducted using group-averaged data. The averaged RDMs for each group, modality and category are presented in Figure 2.

Hierarchical clustering analysis was performed to examine subordinate organizations within categories. Results showed that subordinate structures – kitchenware and tools in HHIs, vegetables and fruits in VFs – were present in children (Figure 3).

To quantify the similarity of hierarchical structures between groups and modalities, cophenetic correlations were computed on the dendrograms derived from the RDMs. These correlations reflect the degree of alignment in clustering structures. As summarized in Table 3, medium-to-high similarities were found in the hierarchical structures of children and adults, both within and across modalities. Notably, cross-modal similarities were slightly higher in adults than in children.

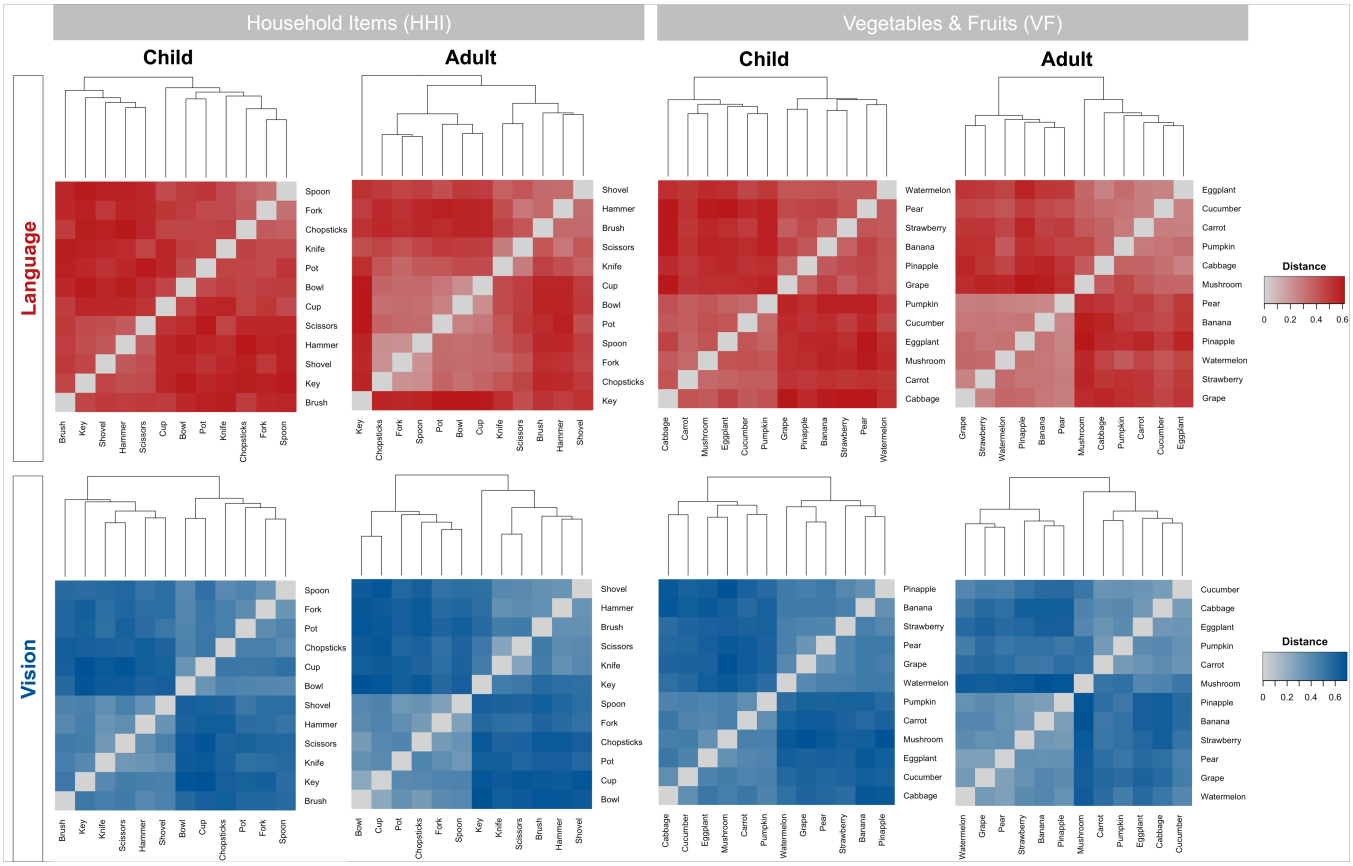

Figure 2: Representational Dissimilarity Matrices (RDMs) computed from the group-averaged pairwise distances in the SpAM task by group (child, adult), modality (language, vision) and category (HHI, VF). Dendrograms indicate results of average linkage hierarchical clustering.

To further assess cross-modal alignment, we computed Mantel tests with 1,000 permutations for each pair of group-averaged RDMs from the two modalities, stratified by group and category. Bootstrapping was used to generate

distributions of correlation coefficients (i.e., alignment scores).

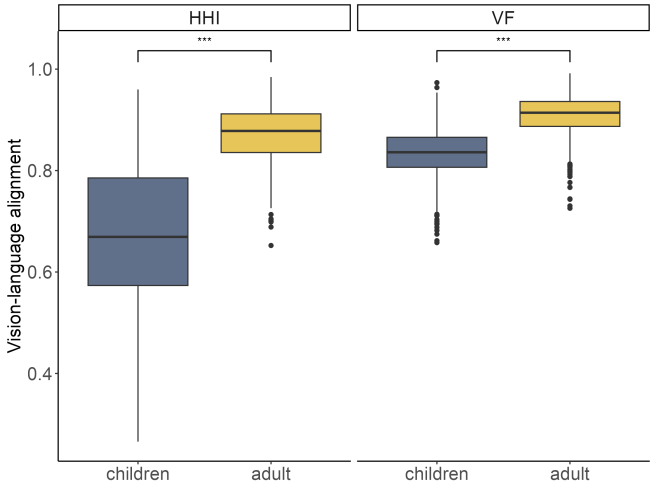

Figure 3: Boxplots of bootstrapped group-level alignment scores across vision-language modalities in children and adults for household items and vegetables & fruits.

Table 3: Results of cophenetic correlations of the hierarchical dendrograms of the RDMs in children and adults. Red numbers indicate cross-modal correlations within group. Blue numbers indicate cross-group correlations within modality. Corrected *p*s < .001.

		Child: vision	Child: language	Adult: vision
HHI	Child: vision	-		
	Child: language	**.61**	-	
	Adult: vision	**.93**	.60	-
	Adult: language	.74	**.57**	**.91**
VF	Child: vision	-		
	Child: language	**.93**	-	
	Adult: vision	**.92**	.88	-
	Adult: language	.91	**.92**	**.96**

The results were largely consistent with the individual-level analysis (Figure 3). Linear regression models were used to assess the effects of group and category on the bootstrapped alignment scores. Significant effects were observed for both predictors. Planned contrasts showed that alignment scores were significantly lower for children compared to adults for both HHIs (t = -.20, p < .001) and VFs (t = -.08, p < .001). Unlike the individual-level analysis, a significant category effect emerged: alignment scores were significantly lower for HHIs compared to VFs for both children (t = -.17, p < .001) and adults (t = -.04, p < .001).

Discussion

In this study, we investigated structural alignment between visual and linguistic presentations of exemplar concepts in children and adults. Participants arranged items presented separately in visual and linguistic modalities on a grid, and the distances between items were used to compute RDMs. These RDMs served as proxies for the structural representations of concepts in psychological space. Notably, all concepts were familiar and readily acquired by the children, ensuring that any developmental differences could not be attributed to concept familiarity. Consistent with our predictions, we found significant lower cross-modal alignment in the similarity structures in children compared to adults. These results were robust across both individual- and group-level analyses.

Developmental Differences in Structural Alignment

Our findings provide evidence that structural refinement continues to occur in the vision and language systems well beyond initial lexical acquisition. While adults demonstrated high alignment across modalities, children's alignment scores were markedly lower, suggesting gradual development in integrating modality-specific structural representations. This likely reflects increasing exposure and semantic detail accumulation of lexical items already in children's vocabularies.

The hierarchical clustering and cophenetic correlation analyses further support this developmental pattern. Although children's hierarchical structures contained adult-like subordinate categories, they were less aligned across modalities. This indicates that while children can organize concepts hierarchically within modalities, cross-modal integration develops more slowly. One explanation is that certain subsections of children's language systems remain under development, particularly for HHI items, which showed weaker organization in the language domain compared to that in the visual modality or to VF items.

Categorical Differences in Structural Alignment

The results revealed a category effect, particularly in the group-level analyses. For both children and adults, VFs showed higher cross-modal alignment than HHIs. This pattern may reflect differences in how concepts of these categories are represented. Natural kind categories like VFs may depend more heavily on shared, modality-independent perceptual features that facilitate cross-modal mapping. Whereas, artifact categories like HHIs might involve more modality-specific functional associations that are harder to integrate. These representational differences were supported by significant variations in psycholinguistic measures, particularly image agreement and name agreement. Such cross-category patterns may explain not only the differential rates of lexical acquisition documented in previous research (Fenson et al., 1994; Opfer & Gelman, 2010) but also the current findings of varied cross-modal alignment.

The reliance on perceptual (e.g., vision) or abstract (e.g., language) systems may also differ across categories during conceptual development. For example, skipgram language models trained on child-directed speech have demonstrated poorer classification performance for HHIs compared to VFs (Asr, Willits, & Jones, 2016), suggesting that language might

provide less robust support to the development of HHI concepts compared to VFs (Li & Zhang, 2025). This aligns with evidence that abstract, functional features, which are particularly relevant for artifact categories, only become prominent in older children, while younger children's representations of artifacts are more constrained by perceptual and mechanical features (Defeyter, Avons, & German, 2007). It is therefore possible that the refinement of artifacts like HHIs may require more developmental time than natural kinds like VFs.

Although we carefully matched stimuli familiarity across categories using adult ratings (Table 1), we acknowledge that children's actual exposure frequency may differ between HHIs and VFs. Children likely have more frequent direct experience with VFs than HHIs in daily life, which could further contribute to the observed alignment differences by providing more opportunities for conceptual refinement. This remains an important consideration for future research to explore systematically.

Implications for the Modality Debate

Most current research assume a unified conceptual system which allows access to meaning regardless of input modalities (Binder & Desai, 2011; Kiefer & Pulvermüller, 2012). Yet there is still ongoing debate over whether conceptual representations are inherently grounded in modality-specific systems or are primarily amodal with modality effects reflecting input processing or task demands (e.g., Calzavarini, 2024). While our study does not directly address this debate, as we explicitly manipulated input modality, our findings suggest two plausible interpretations of the developmental alignment patterns in relation to the debate[1].

First, the visual modality may simply provide stronger or more accessible cues to unified conceptual representations in children, explaining their better performance with visual compared to linguistic stimuli (Table 3). This is also consistent with evidence that visual-perceptual features dominate early conceptual development (Sloutsky, 2010). In this case, we cannot conclude anything about the representational modalities or their alignment. Alternatively, input modalities may differentially activate their corresponding representational systems before converging on amodal representations. Recent MEG evidence supports this hybrid account, showing initial modality-specific activations (~100-200ms) followed by cross-modal convergence (~300-500ms) during conceptual processing (Dirani & Pylkkänen, 2023). While such temporally precise methods help clarify these issues in adults, developmental research faces additional complexity as the conceptual system itself undergoes substantial refinement during childhood.

[1] We thank an anonymous reviewer for suggesting this important point for discussion.

Conclusion

In summary, this study demonstrates that children exhibit weaker structural alignment than adults when accessing familiar concepts across visual and linguistic modalities, with particularly pronounced differences for household items compared to fruits & vegetables. These findings highlight that conceptual development extends well beyond initial item-level learning, involving the progressive alignment of structural mappings across modalities. Future research should continue to investigate how conceptual representations become organized and integrated across development, moving beyond item mappings to examine the complex interplay between different systems.

Acknowledgement

We thank Xin Liang and Jingyi Gu for their assistance with collecting and cleaning the data. Preparation of this paper was supported by The National Social Science Fund of China (23CYY040).

References

Aho, K., Roads, B. D., & Love, B. C. (2022). System alignment supports cross-domain learning and zero-shot generalisation. *Cognition*, *227*, 105200. https://doi.org/10.1016/j.cognition.2022.105200

Aho, K., Roads, B. D., & Love, B. C. (2023). Signatures of cross-modal alignment in children's early concepts. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(42), 1–11. https://doi.org/10.1073/pnas.2309688120

Asr, F. T., Willits, J. A., & Jones, M. N. (2016). Comparing predictive and co-occurrence based models of lexical semantics trained on child-directed speech. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*. Retrieved from https://escholarship.org/uc/item/0g7418hr

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron*, *100*(2), 490–509. https://doi.org/10.1016/j.neuron.2018.10.002

Binder, J., & Desai, R. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536. https://doi.org/10.1016/j.tics.2011.10.001

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Semantic structure in vocabulary knowledge interacts with lexical and sentence processing in infancy. *Child Development*, *87*(6), 1893–1908. https://doi.org/10.1111/cdev.12554

Braginsky, M, Sanchez, A., & Yurovsky, D. (2019). childesr: Accessing the "CHILDES" Database. Retrieved from https://github.com/langcog/childesr

Braginsky, Mika. (2024). wordbankr: Accessing the Wordbank Database. Retrieved from https://langcog.github.io/wordbankr/

Calzavarini, F. (2024). Rethinking modality-specificity in the cognitive neuroscience of concrete word meaning: a position paper. *Language, Cognition and Neuroscience*, *39*(7), 815–837. https://doi.org/10.1080/23273798.2023.2173789

Defeyter, M. A., Avons, S. E., & German, T. C. (2007). Developmental changes in information central to artifact representation: evidence from 'functional fluency' tasks. *Developmental Science*, *10*(5), 538–546. https://doi.org/10.1111/J.1467-7687.2007.00617.X

Dirani, J., & Pylkkänen, L. (2023). The time course of cross-modal representations of conceptual categories. *NeuroImage*, *277*, 120254. https://doi.org/10.1016/j.neuroimage.2023.120254

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., … Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5), i–185. https://doi.org/10.2307/1166093

Gainotti, G., Spinelli, P., Scaricamazza, E., & Marra, C. (2013). The evaluation of sources of knowledge underlying different conceptual categories. *Frontiers in Human Neuroscience*, *7*, 1–12. https://doi.org/10.3389/fnhum.2013.00040

Hills, T. T., & Siew, C. S. Q. (2018). Filling gaps in early word learning. *Nature Human Behaviour*, *2*(9), 622–623. https://doi.org/10.1038/s41562-018-0428-y

Kenett, Y. N., & Faust, M. (2019). A semantic network cartography of the creative mind. *Trends in Cognitive Sciences*, *23*(4), 271–274. https://doi.org/10.1016/J.TICS.2019.01.007

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, *48*(7), 805–825. https://doi.org/10.1016/j.cortex.2011.04.006

León-Villagrá, P., Ehrlich, I., Lucas, C. G., & Buchsbaum, D. (2022). Uncovering children's concepts and conceptual change. In *Proceedings of the 44th Annual Meeting of the Cognitive Science Society*. Retrieved from https://escholarship.org/uc/item/7j67w2nj

Li, L., & Zhang, J. (2025). Tracking conceptual refinement in development: The role of embodied and linguistic systems. *Psyarxiv*.

Luo, K., Zhang, B., Xiao, Y., & Brendan, M. L. (2024). Finding unsupervised alignment of conceptual systems in image-word representations. In *Proceedings of the 46th Annual Meeting of the Cognitive Science Society*. Retrieved from https://escholarship.org/uc/item/7dz6b64q

Lupyan, G., & Lewis, M. (2019). From words-as-mappings to words-as-cues: the role of language in semantic knowledge. *Language, Cognition and Neuroscience*, *34*(10), 1319–1337. https://doi.org/10.1080/23273798.2017.1404114

Mervis, C. B., & Crisafi, M. A. (1982). Order of acquisition of subordinate- , basic- , and superordinate-level categories. *Child Development*, *53*(1), 258–266.

Ni, L., Liu, Y., Yu, W., & Fu, X. (2019). The China image set (CIS): A new set of 551 colored photos with Chinese norms for 12 psycholinguistic variables. *Frontiers in Psychology*, *10*, 2631. https://doi.org/10.3389/FPSYG.2019.02631/BIBTEX

Opfer, J. E., & Gelman, S. A. (2010). Development of the animate – inanimate distinction. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (pp. 213–238). Blackwell.

Roads, B. D., & Love, B. C. (2020). Learning as the unsupervised alignment of conceptual systems. *Nature Machine Intelligence*, *2*(1), 76–82. https://doi.org/10.1038/s42256-019-0132-2

San Roque, L., Norcliffe, E., & Majid, A. (2024). Vision verbs emerge first in English acquisition but touch, not audition, follows second. *Cognitive Science*, *48*(6). https://doi.org/10.1111/cogs.13469

Sizemore, A. E., Karuza, E. A., Giusti, C., & Bassett, D. S. (2018). Knowledge gaps in the early growth of semantic feature networks. *Nature Human Behaviour*, *2*(9), 682–692. https://doi.org/10.1038/s41562-018-0422-4

Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science*, *34*(7), 1244–1286. https://doi.org/10.1111/j.1551-6709.2010.01129.x

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Vales, C., States, S. L., & Fisher, A. V. (2020). Experience-driven semantic differentiation: effects of a naturalistic experience on within- and across-domain differentiation in children. *Child Development*, *91*(3), 733–742. https://doi.org/10.1111/cdev.13369

Vales, C., Stevens, P., & Fisher, A. V. (2020). Lumping and splitting: Developmental changes in the structure of children's semantic networks. *Journal of Experimental Child Psychology*, *199*. https://doi.org/10.1016/j.jecp.2020.104914

Wojcik, E. H., Zettersten, M., & Benitez, V. L. (2022). The map trap: Why and how word learning research should move beyond mapping. *Wiley Interdisciplinary Reviews: Cognitive Science*, *13*(4), e1596. https://doi.org/10.1002/wcs.1596

Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13-and 18-month-olds. *Developmental Psychology*, *30*(4), 553–566.