Gestalt processing in late-sighted children and deep neural networks

Lukas Vogelsang (lvogelsa@mit.edu)
Department of Brain and Cognitive Sciences, MIT
Cambridge, MA 02139, USA

Marin Vogelsang (ozaki@mit.edu)
Department of Brain and Cognitive Sciences, MIT
Cambridge, MA 02139, USA

Priti Gupta (pgupta@projectprakash.org)
Project Prakash, Dr. Shroff's Charity Eye Hospital
Delhi, India

Pragya Shah (paggushah@gmail.com)
Project Prakash, Dr. Shroff's Charity Eye Hospital
Delhi, India

Purva Sethi (purva.sethi15@gmail.com)
Project Prakash, Dr. Shroff's Charity Eye Hospital
Delhi, India

Stutee Narang (narangstutee@gmail.com)
Project Prakash, Dr. Shroff's Charity Eye Hospital
Delhi, India

Suma Ganesh (sumag@sceh.net)
Department of Pediatric Ophthalmology, Dr. Shroff's Charity Eye Hospital
Delhi, India

Pawan Sinha (psinha@mit.edu)
Department of Brain and Cognitive Sciences, MIT
Cambridge, MA 02139, USA

Abstract

Holistic perceptual organization guided by Gestalt principles like similarity in luminance, color, or form is critical for organizing complex visual scenes. Past studies have demonstrated the development of these abilities in infancy. How important is visual experience during these early years for the emergence of Gestalt processing? Does visual deprivation during this period permanently compromise perceptual grouping? Using tasks validated previously with normally-sighted infants, we assessed perceptual grouping from immediately pre-surgery to several months afterward. Late-sighted patients showed persistently poor performance across all Gestalt cues tested, despite longitudinal gains in other visual functions. Computational simulations using deep neural networks trained with typical or atypical developmental acuity trajectories tested the role of early visual degradations characteristic of normal development. Although early degradations enhanced holistic processing in more naturalistic assessments, the benefits for explicit Gestalt processing were very modest. These findings underscore the critical role of early visual experience in developing perceptual organization, reveal important deficits from early deprivation, and suggest new avenues for future computational work.

Keywords: Gestalt principles; perceptual grouping; deep networks; development; early degradations; late-sight onset

Introduction

The Gestalt principles, since their initial formulation (Wertheimer, 1923; Köhler, 1920; Koffka, 1935), have significantly advanced our understanding of how humans organize the complex visual environment into coherent and meaningful percepts (reviewed in Wagemans et al. (2012)). However, the developmental roots and mechanisms underlying this perceptual proficiency are still subject to debate. Two largely independent research efforts have contributed to this discussion: developmental psychology, characterizing the emergence of grouping capabilities during infancy, and computational modeling, testing whether artificial neural networks exhibit Gestalt-like perceptual organization.

Past developmental studies have demonstrated variability in the acquisition timelines of different Gestalt principles, with grouping by luminance similarity observed as early as three months of age (Quinn, Burke, & Rush, 1993) but grouping based on form similarity emerging only by six months (Quinn, Bhatt, Brush, Grimes, & Sharpnack, 2002). Although historically conceived as innate, these findings indicate that at least certain aspects of perceptual grouping depend on visual experience (Quinn & Bhatt, 2006). Yet, it remains unclear whether *early* visual experience is essential for developing robust perceptual grouping abilities and whether deprivation during 'critical periods' permanently impairs holistic visual processing.

Recent computational studies have begun testing Gestalt-like perceptual grouping in deep networks. While initial findings were promising (Kim, Reif, Wattenberg, Bengio, & Mozer, 2021), subsequent research suggests that such capabilities may be limited and vary substantially across architectures, training conditions, and tests (Zhang, Soydaner, Behrad, Koßmann, & Wagemans, 2024; Zhang, Soydaner, Koßmann, Behrad, & Wagemans, 2024; Biscione & Bowers, 2023).

Here, we integrate these two streams of research. We studied a unique group of children who were born blind and gained sight late in childhood following surgical intervention. This population provides unique insights into the role of initial visual experience by enabling assessments immediately upon sight onset. Complementing these empirical findings and examining mechanistic underpinnings, we explored how different developmental trajectories of basic visual functions may influence perceptual grouping capabilities in deep networks.

Human studies

Methods

We adapted paradigms previously validated to assess Gestalt-like perceptual grouping in infants (Quinn et al., 1993, 2002; Quinn, Bhatt, & Hayden, 2008; Hayden, Bhatt, & Quinn, 2006). To test grouping by similarity in luminance and form, participants viewed square arrays arranged in a uniform grid, where spatial arrangement alone provided no grouping cues. Instead, identical luminance or figural form assigned to alternating rows or columns induced grouping (Quinn et al., 1993, 2002). Participants indicated perceived vertical or horizontal organization via a 2AFC response. Proximity-based grouping was assessed with a 3x3 grid unequally spaced along one dimension (Quinn et al., 2008). Connectedness-based grouping was tested by figures separated into connected or disconnected components (Hayden et al., 2006). Participants completed 10 randomized trials for each of six conditions (illustrated in Figure 1A) without time constraints.

Our participants comprised ten children with early-onset blindness identified through pediatric ophthalmic screening programs in rural villages. Parental reports and ophthalmic examinations indicated severe early visual impairment, with participants categorized as having 'Profound visual impairment' (20/500–20/1000) or 'Light perception/projection' ($<$20/1000) pre-operatively. Testing occurred pre-operatively and longitudinally up to one month post-operatively, with a subset (n=5) additionally tested about four months post-surgery.

Results

Figure 1B shows pre-operative and one-month post-operative results for the full participant group (n=10). No clear improvements were observed, with the proximity condition, despite negligible longitudinal gains, yielding the highest absolute performance. Figure 1C shows similarly low performance for the subset of late-sighted children (n=5) tested approximately four months post-operatively. By this time, even typically developing infants demonstrate sensitivity to some Gestalt cues.

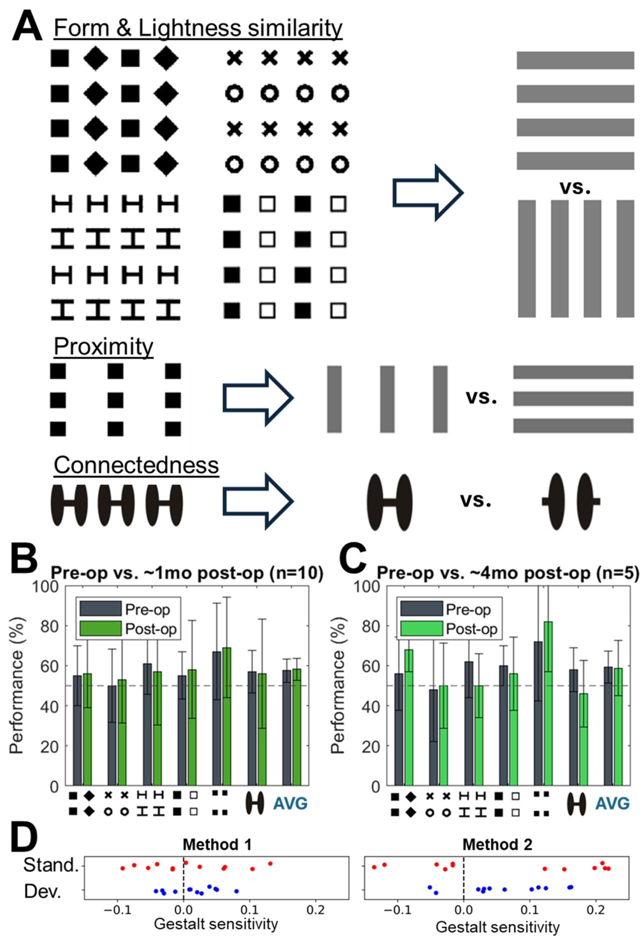

Figure 1: A. The 6 types of Gestalt displays used. B&C. Performance of late-sighted children pre-op vs. 1-month (B) / 4-months (C) post-op. D. Gestalt sensitivity for both models (standard vs. developmental) and both methods, with 2 stimulus versions used for each of the 6 Gestalt cues.

Computational studies

Methods

Next, we explored how developmental trajectories of basic visual functions may influence perceptual grouping abilities computationally. Past work proposed that holistic processing may emerge adaptively from poor visual acuity early in visual development, although the exact benefits vary (Vogelsang et al., 2018; Jang & Tong, 2021; Yoshihara, Fukiage, & Nishida, 2023). Late-sighted individuals lack this trajectory, commencing vision with better-than-neonatal acuity. To this end, we trained the AlexNet (Krizhevsky, Sutskever, & Hinton, 2012) on ImageNet (Deng et al., 2009) using two training trajectories: a 'standard' one trained only with high-resolution images, and a 'developmental' one trained initially with blurred images and transitioning midway through training to clear images.

To quantify Gestalt-like processing, we adapted the methodology from Kim et al. (2021) and extracted representations from the penultimate network layer. Specifically, we computed the similarity of unit activations between a test stimulus and each of the two alternative forced-choice options, quantifying Gestalt sensitivity as the similarity advantage for the correct vs. incorrect structure. In 'Method 1', we used the actual horizontal vs. vertical bar stimuli from the experiment; in 'Method 2', we used systematic deletion of the middle 50% of elements along the horizontal vs. vertical axis. Per Gestalt cue and method, two different stimulus versions were tested.

Results

Figure 1D depicts the Gestalt sensitivity scores for our two models. While, across both methods, the developmental model yields greater-than-zero scores for more stimuli, the standard model's larger variance leads to several stimuli having higher scores. Overall, both models exhibit fairly low absolute Gestalt scores, with relative model differences, too, being modest. Interestingly, this contrasts with more naturalistic assessments of holistic processing, such as global-shape biases, based on the methodology by Geirhos et al. (2018), for which the developmental model shows clear benefits (with correct decisions driven by shape by 46% in the standard model but 66% in the developmental model).

Discussion

Behaviorally, our findings extend past evidence that late-sighted patients exhibit long-lasting deficits in holistic processing (De Heering & Maurer, 2014; Geldart, Mondloch, Maurer, De Schonen, & Brent, 2002; Vogelsang et al., 2018), now including the important domain of Gestalt grouping. It remains to be seen whether some aspects of Gestalt processing can be acquired with more extensive visual experience.

Computationally, our results are more nuanced. Although prior work suggests adaptive advantages for initially poor acuity in promoting holistic processing, and even though our developmental model showed an increased shape bias, benefits for explicit Gestalt stimuli were very modest. This divergence could indicate that explicit Gestalt grouping is more challenging to achieve or, alternatively, that our out-of-distribution stimuli and analysis methodology lack sufficient sensitivity for uncovering developmental effects. Moving forward, refining stimuli and analytical methods could further elucidate computational analogs of Gestalt-like organization. Notwithstanding these limitations, our approach exemplifies how computational modeling can help advance developmental science and how, in turn, developmental insights can provide inspiration for more human-like machine learning models.

Acknowledgements

This work has been supported by NIH grant R01EY020517 to Pawan Sinha. Lukas Vogelsang is supported by a grant from the Simons Foundation International to the Simons Center for the Social Brain at MIT. Marin Vogelsang is supported by the Japan Society for the Promotion of Science (JSPS), Overseas Research Fellowship and the Yamada Science Foundation.

References

Biscione, V., & Bowers, J. S. (2023). Mixed evidence for gestalt grouping in deep neural networks. *Computational Brain & Behavior*, *6*(3), 438–456.

De Heering, A., & Maurer, D. (2014). Face memory deficits in patients deprived of early visual input by bilateral congenital cataracts. *Developmental Psychobiology*, *56*(1), 96–108.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 ieee conference on computer vision and pattern recognition* (pp. 248–255).

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2018). Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. In *International conference on learning representations.*

Geldart, S., Mondloch, C. J., Maurer, D., De Schonen, S., & Brent, H. P. (2002). The effect of early visual deprivation on the development of face processing. *Developmental Science*, *5*(4), 490–501.

Hayden, A., Bhatt, R. S., & Quinn, P. C. (2006). Infants' sensitivity to uniform connectedness as a cue for perceptual organization. *Psychonomic Bulletin & Review*, *13*, 257–261.

Jang, H., & Tong, F. (2021). Convolutional neural networks trained with a developmental sequence of blurry to clear images reveal core differences between face and object processing. *Journal of vision*, *21*(12), 6–6.

Kim, B., Reif, E., Wattenberg, M., Bengio, S., & Mozer, M. C. (2021). Neural networks trained on natural scenes exhibit gestalt closure. *Computational Brain & Behavior*, *4*(3), 251–263.

Koffka, K. (1935). *Principles of gestalt psychology*. London, U.K.: Lund Humphries.

Köhler, W. (1920). *Die physischen gestalten in ruhe und im stationären zustand: Eine naturphilosophische untersuchung*. Vieweg+ Teubner Verlag.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

Quinn, P. C., & Bhatt, R. S. (2006). Are some gestalt principles deployed more readily than others during early development? the case of lightness versus form similarity. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1221.

Quinn, P. C., Bhatt, R. S., Brush, D., Grimes, A., & Sharpnack, H. (2002). Development of form similarity as a gestalt grouping principle in infancy. *Psychological science*, *13*(4), 320–328.

Quinn, P. C., Bhatt, R. S., & Hayden, A. (2008). Young infants readily use proximity to organize visual pattern information. *Acta Psychologica*, *127*(2), 289–298.

Quinn, P. C., Burke, S., & Rush, A. (1993). Part—whole perception in early infancy: Evidence for perceptual grouping produced by lightness similarity. *Infant Behavior and Development*, *16*(1), 19–42.

Vogelsang, L., Gilad-Gutnick, S., Ehrenberg, E., Yonas, A., Diamond, S., Held, R., & Sinha, P. (2018). Potential downside of high initial visual acuity. *Proceedings of the National Academy of Sciences*, *115*(44), 11333–11338.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & Von der Heydt, R. (2012). A century of gestalt psychology in visual perception: I. perceptual grouping and figure–ground organization. *Psychological bulletin*, *138*(6), 1172.

Wertheimer, M. (1923). Laws of organization in perceptual forms.

Yoshihara, S., Fukiage, T., & Nishida, S. (2023). Does training with blurred images bring convolutional neural networks closer to humans with respect to robust object recognition and internal representations? *Frontiers in Psychology*, *14*, 1047694.

Zhang, Y., Soydaner, D., Behrad, F., Koßmann, L., & Wagemans, J. (2024). Investigating the gestalt principle of closure in deep convolutional neural networks. *arXiv preprint arXiv:2411.00627*.

Zhang, Y., Soydaner, D., Koßmann, L., Behrad, F., & Wagemans, J. (2024). Finding closure: A closer look at the gestalt law of closure in convolutional neural networks. *arXiv preprint arXiv:2408.12460*.