Visual production & recognition

https://doi.org/10.1038/s44159-024-00321-0

Shaping vision through drawing

Human communication, often framed in the context of natural language, consists of two tightly coupled processes — production and recognition. Whereas the latter is a regular fixture in the study of vision, rigorous studies on visual production have come into the limelight only in the past few decades. The study of visual communication applies the production–comprehension framework to investigate how people convey ideas through visual conduits such as graphs, diagrams, symbols and drawings. Freehand drawings, in particular, constitute a historically and culturally pervasive, flexible and accessible medium for externalizing and preserving thoughts.

A key characteristic of many freehand drawings (such as ancient cave art or a toddler's sketches) is that they are abstract renderings that do not strictly adhere to the specific visual features of the target object concept. Yet, despite this visual divergence from real world referents, people are able to recognize drawings and use them for communication, planning and pedagogy.

Fan and colleagues proposed that the human visuo-semantic system solves this problem of visual abstraction by recruiting the same representations for drawing production that it uses for recognizing drawings and real-world objects. Using a deep neural network model trained on photographs for categorization, they established that photographs and drawings of the same object classes (for example, drawings and photographs of fish) were represented as more similar than those belonging to different classes (for example, drawings of fish and photographs of cars). This finding, coupled with the fact that this model was consistent with human categorization of sketches, led the authors to adopt the model's representations as a proxy for human visual representations. The authors then showed that, although a drawing training task led people to produce drawings that were more recognizable relative to drawings made by those who did not undergo the training, these improvements in recognizability did not extend to drawings of objects in different categories or objects in the same category that the participants did not practice. Thus, production practice influenced recognizability in a precise, targeted manner.

> "action and perception coupled together jointly shape mental representations"

Indeed, a fine-grained look at the neural network representations of the drawings made during this task revealed that improvements in recognizability were primarily driven by a 'sharpening' of object-specific visual features used to depict the object in drawing form. That is, the mental representations that people were using to depict the object (functionally captured by the neural network representations) became more aligned with the representations used for recognition.

Last, the authors showed that the active task of drawing production was critical for these improvements. Participants who saw only stroke-by-stroke recreations of drawings (versus making the drawings themselves) failed to show any improvements in a perceptual discrimination task.

The notion that action and perception coupled together jointly shape mental representations has served as a north star in steering my own research program and is foundational to my understanding of human semantic knowledge. Part of what made this paper possible was the development of image-computable models of visual recognition. Continued advances in models of human-like sketch production offer exciting new avenues for linking perception to action by building mechanistic accounts of how drawing production shapes visual representations couched in computational cognitive models.

Kushin Mukherjee [1,2]

[1]Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA.
[2]Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI, USA.
e-mail: kushinm11@gmail.com

Competing interests
The author declares no competing interests.

Original article: Fan, J. E., Yamins, D. L. & Turk-Browne, N. B. Common object representations for visual production and recognition. *Cogn. Sci.* **42**, 2670–2698 (2018)

Related articles: Patterson, K., Nestor, P. J. & Rogers, T. T. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* **8**, 976–987 (2007); Vinker, Y., Alaluf, Y., Cohen-Or, D. & Shamir, A. Clipascene: scene sketching with different types and levels of abstraction. In *Proc. IEEE/CVF International Conference on Computer Vision* 4146–4156 (CVF, 2023)