Perceptual expertise and object recognition
An explanatory task for modularists and antimodularists

Aleksandra Mroczko-Wąsowicz[a] (aleksandra.mroczko-wasowicz@uw.edu.pl)

Abstract

Dustin Stokes's book contributes to one of the continuing debates in empirically informed philosophy of mind and cognitive sciences which concerns the relation between thought and perception. The book sheds new light on such questions as: whether vision is modular, informationally encapsulated, and thus cognitively impenetrable or rather the opposite – whether it is malleable and sensitive to further improvements by cognitive states. Stokes supports the latter by referring to empirical evidence on perceptual expertise. Proponents of the modular and malleable architectures of the mind offer different explanations of the phenomena involved in perceptual expertise, viz. object identification and categorization. Interestingly, both views assume some kind of automaticity of the recognitional capacities for identifying and categorizing objects. In this article, I examine the influence of perceptual expertise on object recognition and how the seeming automaticity of object recognition may be approached from the modularist and antimodularist (malleabilist) perspectives.

Keywords

Perceptual expertise · Object recognition · Automaticity · Modularity

This article is part of a symposium on Dustin Stokes's book "Thinking and Perceiving" (Routledge 2021), edited by Regina Fabry and Sascha Benjamin Fink.

Dustin Stokes's book sheds new light on processes that constitute the human mind, the way they interact, and enable us to make contact with the world. He approaches the relation between thought and perception by dealing with such questions as: whether vision is modular, informationally encapsulated, and thus cognitively impenetrable or rather the opposite – whether it is malleable and sensitive to further improvements by cognitive states. Stokes supports the latter by appealing to empirical research on perceptual expertise (Bukach et al., 2006; Drew et al., 2013; Kundel et al., 2007; Scott, 2011) and his own investigations of this phenomenon (Ivy et al., 2021, 2023; Stokes, 2021a).

---

[a]University of Warsaw, Faculty of Philosophy.

Expertise involves experience and training specific to a relevant domain such as radiology, ornithology, or fingerprint examining. Proponents of the modular and malleable architectures of the mind offer different explanations of phenomena involved in perceptual expertise, viz. object categorization and identification. Modularists interpret it as the capacity for visual object recognition – a post-perceptual cognitive process of late vision. For Stokes, perceptual expertise, including expert recognition, is a genuinely perceptual phenomenon, which can be sensitive to cognitive influences and as such exhibits both a perceptual and cognitive achievement (Stokes, 2021a, 2021b).

Capacities for object recognition are capacities to either categorize a perceived object as belonging to a conceptual category or to identify it as being a specific individual (Abid, 2021). The two processes can be seen as different stages of the recognition of an item, with categorization being temporally more efficient and occurring before identification. There is wide agreement that at least some of these processes operate automatically (Dell'acqua & Job, 1998; Mroczko-Wąsowicz, & Anaya 2022; Serre et al., 2007). A process is automatic, as opposed to voluntary or intentional, when it is not under the subject's conscious control (Moors & De Houwer, 2006; Papineau, 2013, p. 177). Empirical studies have determined that recognition can occur as quickly as within 200-300 ms after stimulus detection (DiCarlo et al., 2012; Mohr et al., 2018). This short time frame excludes any possibility for subjects to exercise conscious control over the unfolding of the process.

Stokes claims that recognition by experts differs from that by novices. This is because expert perception is cognitively enhanced as a result of their domain-specific concept-rich cognitive learning and unconscious impact of perceptual learning (visual memories) on attention. As a result, they do not only know better but see better. The empirical findings Stokes (2021b, p. 193) refers to suggest that in expert recognition, the memorized visual information affects saccadic patterns so that experts are less distracted by features of a stimulus that have no consequence for recognizing a specific category of objects. Stokes discusses eye-tracking studies which have shown that experts such as radiologists make fewer but longer saccades and fixate less on features of a radiogram that are diagnostically irrelevant for identifying an abnormality (Drew et al., 2013; Kundel et al., 2007). In consequence, decreased visual distraction leads to enhanced sensitivity to category-specific information. Experts develop visual object representations of the relevant kind more rapidly and see relevant objects more accurately than laypersons. This is why it is concluded that experts display differences in eye movement patterns, enjoy an advantage for a specific category of objects in visual short-term memory, and exhibit automatic successful performance (Stokes, 2023).

In a similar vein, Stokes and colleagues argue in their own experimental work that visual expertise is more than meets the eye (Ivy et al., 2023). They examined holistic visual processing (HVP) – a behavioral marker of a visual-expert search strategy, which shows that experts are able to process information from a larger

region of space with a more focused gaze pattern. HVP turns out to be transferable across domains but supporting reduced search time and greater accuracy only within an expert's particular domain of expertise. This led the researchers to conclude that visual search success does not depend exclusively on the occurrence of HVP, but also on the explicit knowledge of an expert's domain, including their knowledge how to search and where to search.

The seeming automaticity of object recognition may be an outcome of the interplay between these factors, because recognition is a complex phenomenon consisting of perception, concepts, and associated perceptual memories of categories or particular items. Commenting on behavioral and phenomenological aspects of expert recognitional capacity, Stokes emphasizes its instantaneousness:

> The first strand of behavioural evidence concerns "automaticity". Not only do experts more rapidly perform categorizations or other forms of recognition (...), but they do so in ways that they often cannot carefully describe. (...) expert radiologists often report a sense that there is something anomalous in a medical image before they can point to the anomaly. (...) [they] report that the relevant object or feature is "highly salient" or just "pops out". (...) These reports and the speed of performance suggest that the expert expends little or no deliberate cognitive effort and that her performance is non-inferential. (Stokes, 2021b, p. 152)

On the other side, theorists aligning toward the modularist view (Fodor, 1983; Marr, 1982; Pylyshyn, 1999) may explain visual object recognition differently. This explanation also involves some aspects of recognition's automaticity, but this is limited to the early sensory component of recognition. Accordingly, visual recognition includes automatic early sensory processing delivering new visual representations of basic sensory properties for purposes of later categorization and identification. This means that the new visual representations are compared with stored visual memories. In consequence, feedback from visual memories helps to search the newly formed sensory representations for visual matches. Even quick comparisons, which lead to finding recognitional matches between novel visual representations and those stored in memory, are large enough to accommodate attention allocation and to be classified by modularists as late vision or post-perceptual processes (Raftopoulos, 2011).

Modularists acknowledge that late vision is cognitively penetrated and involves modulation of processing by cognitively driven attention (Pylyshyn, 2003). By classifying visual object recognition as a late vision phenomenon, they do not mean that object recognition enhanced by perceptual expertise is an *additional* indicator of the malleability of the mind. For modularists, perceptual expertise does not make object recognition more sensitive to cognitive influences than it is already the case for regular (non-expert) recognition. On their view, perceptual expertise makes the low-level component of object recognition – early

visual sensory processing – more fluent and facilitates perceptual discriminations of fine-grained subordinate categories. Modularists emphasize that outputs of early vision result from unconscious, mandatory, fast, and automatic modular processes. As such, early visual representations do not alone determine late vision phenomena like object recognition. Interestingly, although automaticity in the modularist view is meant in the strong sense of operating pre-attentively, it can be considered to be a graded matter, similarly to modularity itself (Deroy, 2014; Drayson, 2017; Fodor, 1983; Mroczko-Wąsowicz, 2022).

In this commentary, I have examined how object recognition and its automaticity may be approached from the modularist and malleabilist perspectives. Although some relevant questions regarding the status of late vision remain open for further research, it seems that non-modular approaches to object recognition may enjoy more advantages. This is because such approaches are capable of accommodating twofold explanations of the rapidity of expert recognition: (1) bottom-up explanation compatible with the modularist suggestion concerning the extraordinary fluency of discriminatory processes in the early sensory component of object recognition, and (2) the malleabilist proposal concerning the top-down penetrating impact of expertise-related higher cognitive states on perceptual processing and perceptual phenomenology.

Even if one agrees with modularists that the influence of expertise on object recognition can be interpreted as cognitive effects on post-perceptual processes, this does not have to be disruptive to Stokes's proposal. It seems his proposal is capable of acknowledging that cognitive contact with the world can have an effect not only on perceptual contact with the world but also on related hybrid phenomena such as recognition (Stokes, 2021b, pp. 3–7). The main thesis of the book, namely the cognitive improvement of perception in cases of top-down effects of expertise on recognition, remains untouched by the pertinent warning by Firestone & Scholl (2016, pp. 15–17) who suggest not to confuse two integral but separable constituents of recognition: perception and memory. Carefully distinguishing these constituents is essential, because cognitive effects on back-end memory have no implications for front-end perception. *Thinking and Perceiving* avoids this pitfall.

Why does explaining expert recognition matter? Providing a satisfactory account of the smoothness of perceptual recognition of objects for which we serve as experts would not only inform debates in philosophy of mind/perception and the cognitive sciences but also furnish us with a better understanding of the process that is ubiquitous in ordinary life and that is central in efficient interactions with objects in our environment (see Mroczko-Wąsowicz & Grush, 2023). The reason is that we all are experts in perceiving and recognizing some kinds of objects (Mroczko-Wąsowicz et al., 2023).

References

Abid, G. (2021). Recognition and the perception–cognition divide. *Mind and Language*, *37*((5)), 770–789. https://doi.org/10.1111/mila.12362

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: The power of an expertise framework. *Trends in Cognitive Sciences*, *10*(4), 159–166. https://doi.org/10.1016/j.tics.2006.02.004

Dell'acqua, R., & Job, R. (1998). Is object recognition automatic? *Psychonomic Bulletin & Review*, *5*, 496–503. https://doi.org/10.3758/BF03208828

Deroy, O. (2014). Modularity of perception. In M. Matthen (Ed.), *Oxford handbook of philosophy of perception* (pp. 755–778). Oxford University Press. https://doi.org/10.1148/rg.331125023

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition. *Neuron*, *73*, 415–434. https://doi.org/10.1016/j.neuron.2012.01.010

Drayson, Z. (2017). Modularity and the predictive mind. In T. M. & W. Wiese (Ed.), *Philosophy and predictive processing* (pp. 1–12). Frankfurt am Main: MIND Group. https://doi.org/DOI: 10.15502/9783958573130

Drew, T., Evans, K., Vő, M. L.-H., Jacobson, F. L., & Wolfe, J. M. (2013). Informatics in radiology: What can you see in a single glance and how might this guide visual search in medical images? *Radiographics*, *33*, 263–274. https://doi.org/10.1148/rg.331125023

Firestone, C., & Scholl, B. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, *39E229*, E229. https://doi.org/10.1017/S0140525X15000965

Fodor, J. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Ivy, S., Rohovit, T., Lavelle, M., Padilla, L., Stefanucci, J., Stokes, D., & Drew, T. (2021). Through the eyes of the expert: Evaluating holistic processing in architects through gaze-contingent viewing. *Psychonomic Bulletin & Review*, *28*(3), 870–878. https://doi.org/10.3758/s13423-020-01858-w

Ivy, S., Rohovit, T., Stefanucci, J., Stokes, D., Mills, M., & Drew, T. (2023). Visual expertise is more than meets the eye: An examination of holistic visual processing in radiologists and architects. *Journal of Medical Imaging*, *10*(1), 1–15. https://doi.org/10.1117/1.JMI.10.1.015501

Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic componentof image perception in mammogram interpretation: Gaze-tracking study. *Radiology*, *242*, 396–402. https://doi.org/10.1148/radiol.2422051997

Marr, D. (1982). *Vision: A computational investigation into human representation and processing of visual information*. San Francisco: Freeman.

Mohr, S., Wang, A., & Engell, A. D. (2018). Early identity recognition of familiar faces is not dependent on holistic processing. *Social Cognitive and Affective Neuroscience*, *13*(10), 1019–1027. https://doi.org/10.1093/scan/nsy079

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326. https://doi.org/10.1037/0033-2909.132.2.297

Mroczko-Wąsowicz, A. (2022). Modularity. In B. D. Young & C. Dicey Jennings (Eds.), *Mind, cognition, and neuroscience: A philosophical introduction* (pp. 149–163). New York: Routledge Press.

Mroczko-Wąsowicz, A., & Anaya, A (2022). How to Explain the Automaticity of Object Recognition. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni. *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

Mroczko-Wąsowicz, A., & Grush, R. (Eds.). (2023). *Sensory individuals: Unimodal and multimodal perspectives*. Oxford: Oxford University Press.

Mroczko-Wąsowicz, A., O'Callaghan, C., Cohen, J., Scholl, B., & Kellman, P. (2023). Advances in the study of visual and multisensory objects. *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

Papineau, D. (2013). In the zone. *Royal Institute of Philosophy Supplements*, *73*, 175–196. https://doi.org/10.1017/S1358246113000325

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*(3), 341–365. https://doi.org/10.1017/s0140525x99002022

Pylyshyn, Z. (2003). *Seeing and visualizing: It's not what you think*. Cambridge, MA: MIT Press.

Raftopoulos, A. (2011). Late vision: Processes and epistemic status. *Frontiers in Psychology*, *2*(382), 382. https://doi.org/10.3389/fpsyg.2011.00382

Scott, L. S. (2011). Face perception and perceptual expertise in adult and developmental populations. In G. R. Rhodes, J. Haxby, M. Johnson, & A. Calder (Eds.), *Oxford handbook of face perception* (pp. 195–214). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199559053.013.0011

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(15), 6424–6429. https://doi.org/10.1073/pnas.0700622104

Stokes, D. (2021a). On perceptual expertise. *Mind and Language*, *36*(2), 241–263. https://doi.org/10.1111/mila.12270

Stokes, D. (2021b). *Thinking and perceiving: On the malleability of the mind.* Routledge/Taylor & Francis Group.

Stokes, D. (2023). Précis of thinking and perceiving. *Philosophy and the Mind Sciences*, 10.