Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Psychology

Reading scenes: how scene grammar guides attention and aids perception in real-world environments

Melissa Le-Hoa Võ[1], Sage EP Boettcher[2] and Dejan Draschkow[1]

Despite many recent technical advances, the human efficacy of naturalistic scene processing is still unparalleled. What guides attention in real world environments? How does scene context affect object search and classification? And how are the many predictions we have with regards to our visual environment structured? Here, we review the latest findings that speak to these questions, ranging from traditional psychophysics, eye movement, and neurocognitive studies in the lab to experiments in virtual reality (VR) and the real world. The growing interest and grasp of a scene's inner workings are enriching our understanding of the efficiency of scene perception and could inspire new technological and clinical advances.

Addresses
[1] Department of Psychology, Johann Wolfgang-Goethe-Universität, Frankfurt, Germany
[2] Department of Experimental Psychology, University of Oxford, Oxford, England, United Kingdom

Corresponding author:
Võ, Melissa Le-Hoa (mlvo@psych.uni-frankfurt.de)
URL: http://www.SceneGrammarLab.com (M.-H. Võ).

Current Opinion in Psychology 2019, **29**:205–210

This review comes from a themed issue on **Attention and perception**

Edited by **Sarah Shomstein**, **Andrew Leber** and **Joy Geng**

https://doi.org/10.1016/j.copsyc.2019.03.009

Perception is much more than meets the eye. Looking at this paper, you likely have no problem deciphering these words. You also know that the somewhat blurry item on your right is your coffee mug and (probably!) not a roll of toilet paper. While we easily accomplish seeing and interacting with our environment, the underlying computations of object perception and attentional deployment are far from trivial. You start realizing how complex human visual cognition is when you try to teach a robot to see, let alone interact with its surroundings. What is easy for an infant (e.g. recognizing a teddy from different viewpoints or localizing it even when hiding underneath a blanket), still poses a major stumbling block for sophisticated computer vision algorithms (for a review see Ref. [1]). In this paper, we will review what guides attention in real-world scenes, what makes object perception so efficient, and how predictions in scenes might form a hierarchy to benefit object perception and search. Importantly, we will argue that we have learned the rules of the world like we learn the rules of our mother tongue, that is, without explicit training, but rather through constant interactions with our world. These sets of rules — one's 'scene grammar' — are the key ingredients to efficient object perception and search.

What guides attention in the real world?

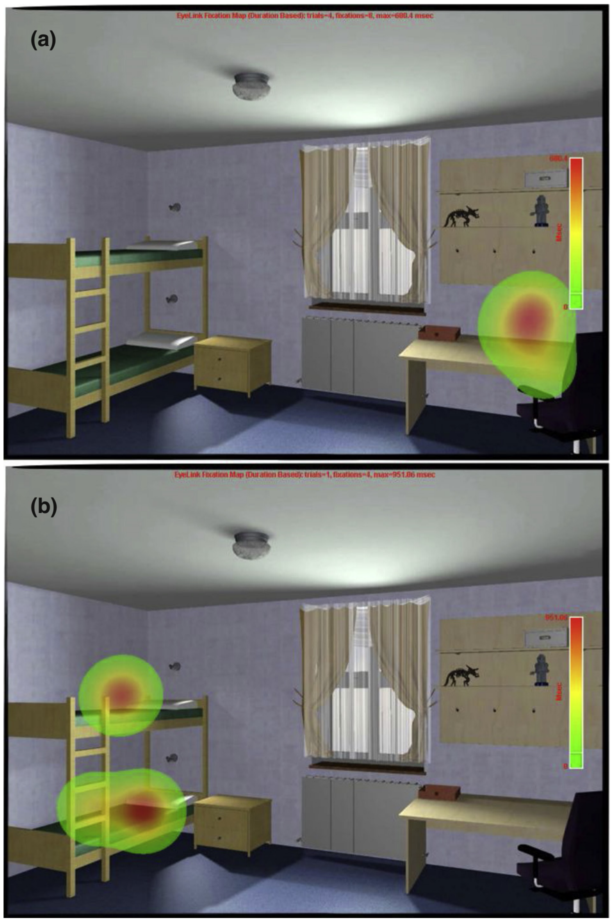
The *perception* of naturalistic scenes is an incredibly fascinating topic of study. Not only can we see the forest without representing the trees [2], but neural signals distinguish both basic-level categories and global properties of scenes within 100 ms [3]. Having access to the gist of a scene within the blink of an eye [4,5] allows us to use this readily available information by activating stored knowledge regarding the typical composition of scenes. With such an instantaneous and efficient perceptual processing, what is it that subsequently guides attention within such complex scenes?

Saliency models have shown that purely bottom-up feature contrasts can successfully predict the allocation of attention in scenes [6], especially when objects as mid-level features are taken into account [7]. Also recent advances in object recognition via deep neural networks have shown that gaze can be predicted purely by identifying objects within a scene [8]. MASC, a model of attention inspired by the superior colliculus captures neurophysiological constraints on saccade programming and is able to predict the fixation locations of observers freely viewing naturalistic scenes and performing search tasks [9•], while the LATEST model can not only explain *where* observers look, but also *when* [10•]. In all of the above, low-level visual features maintain a central role in the modelling approach. However, we often look for things outside of our visual field or hidden from view, which renders saliency less important than our knowledge about objects that are not yet visible [11]. In addition, object functions can also guide attention in scenes [12].

According to the 'cognitive guidance theory', meaning is the main guiding factor within complex, naturalistic scenes directing attention to scene regions that are semantically informative and cognitively relevant in the situation at hand [11,13,14]. The main assumption of this theory is that a first 'parse' of the scene generates a 'flat' landscape of potential attentional targets that is independent of local image saliency. Only knowledge representations then assign attentional priority to targets based on their meaning.

This dominance of top–down knowledge over bottom–up saliency has been demonstrated many times [11,15], for example, when a large, salient toothbrush was missed by human observers simply because it was not predicted by top–down expectations [16]. The strength of predictions also becomes apparent when searching for objects currently absent from a scene as shown in Figure 1. Observers deployed their attention — here measured through eye movements — differentially depending on the search target (‘laptop’ versus ‘teddy’), despite an unchanging scene that did not provide bottom–up features of the targets. When moving around in real 3D environments, cognitive relevance together with our interactions with objects play a strong role in how we memorize and shift attention in our world [17]. It is, therefore, essential to further our understanding of the modulators of visual processing of objects in scenes.

Figure 1



Current Opinion in Psychology

Fixation distributions of an observer looking for a ‘laptop’ in **(a)** and a ‘teddy’ in **(b)** within the identical scene and without target features to guide attention.

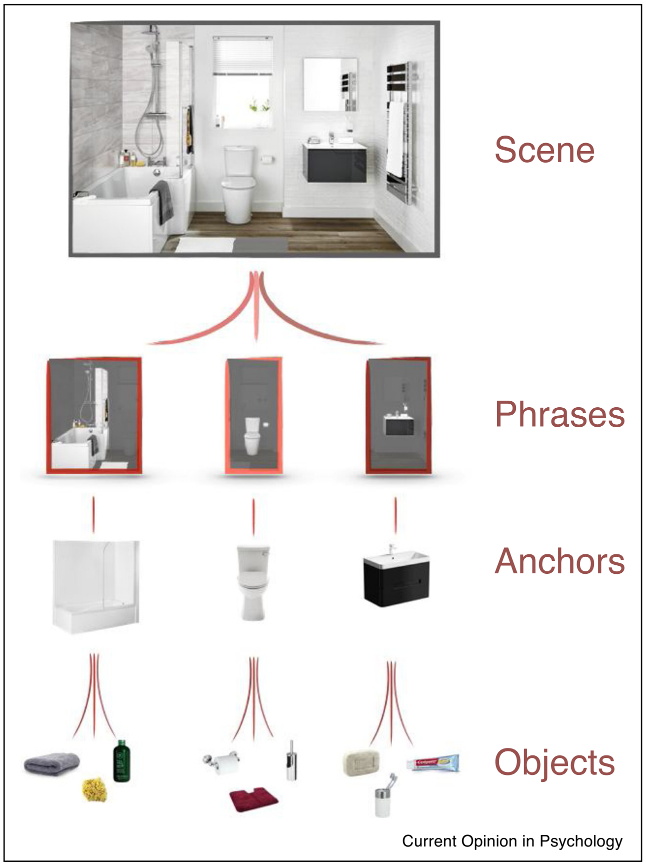
Scene context effects on object perception and search

A fascinating asset of studying scene perception is that scenes are complex, but at the same time are organized according to a set of rules. More specifically, objects in scenes — like words in sentences — seem constrained by a ‘grammar’ that we have implicitly learned and that allows us to efficiently understand scenes, recognize the objects embedded within them, and guide goal-directed behavior [18]. Objects, for example, do not hover in mid-air and hardly ever appear in isolation. Instead they tend to rest on surfaces and are often encountered in similar, repeating surroundings. This scene grammar provides strong priors regarding *what* objects tend to be *where* within certain scenes. In language, semantics refer to the study of the relationship between words, while the study of syntax investigates principles and rules determining the structure of a sentence.

Accordingly, the terms ‘semantics’ and ‘syntax’ have been used to describe object-scene relationships [19•,20]. For example, a semantic violation refers to an object that does not fit the overall meaning of the scene category (e.g. a fire-hydrant in a kitchen), while an object that is placed in a position that is not predicted by the local structure of the scene (e.g. a toaster in the kitchen sink) is considered a syntactic violation. Which of these rules are innate and which learned, is not entirely clear. It is unlikely that one is born with knowledge of where the toothbrush is to be found, but as Eleanor Gibson nicely put it: “animals needn’t learn to perceive; rather, they perceive to learn” [21,22]. In any case, scene grammar provides strong priors regarding what objects tend to be where within certain scenes.

Inconsistent scene contexts, have shown not only to impede search (e.g. [23,24,25]; for a review see [26]), but also disrupt the correct identification of objects embedded within them [19•,27,28]. In addition, when viewing images that contain such violations, observers tend to look longer at inconsistent compared to consistent objects reflecting increased attentional demands [20,29]. This is true even when the task does not involve processing the scene or its embedded objects, for instance when searching for overlaid Ts amongst Ls [30]. Moreover, when asked to classify object or scene words overlaid on to be ignored images, performance for both object and scene words is diminished when the image is incongruent with the word [31]. These findings imply a powerful, automatic processing of scenes, while the question regarding the necessity of attention for correct scene categorization is still under debate [28,32,33]. Note that the central points of this paper are in our view largely independent of the ongoing debate on whether

Figure 2



Current Opinion in Psychology

Schematic hierarchy of a bathroom scene with three phrases consisting of one anchor each (e.g. a shower, a toilet and a sink) that predict the locations of other objects (e.g. the shower predicting the

contextual effects are reflexive or whether they constitute a form of cognitive penetrability (for an in-depth discussion see Ref. [34]).

Similar to language processing, objects semantically inconsistent with regard to their global scene context elicit an N400 ERP response that signifies semantic integration costs [35–38]. In addition, an earlier N300 response is hypothesized to reflect pre-semantic perceptual processes [37,39], but while an early appearance of high-level, context-sensitive processes in visual cognition is undoubted, the existence of two separable neurocognitive processes has been recently questioned [40]. Finally, as in language syntactically/structurally inconsistent objects and events have shown to evoke a left-lateralized anterior negativities (LANs) and/or P600 responses that are clearly different from responses to semantic inconsistencies [36,41•,42]. Thus, different object-scene inconsistencies elicit differential brain responses.

Note that the use of the terms semantics and syntax should not imply that scene and linguistic processing are viewed as being the same, nor should the two categories be seen as easily separable, but instead they likely lie on a continuum. With growing data sets and annotation possibilities this continuum will be mapped out in the future. A large portion of what we know about language processing has been established by using rich and exhaustively annotated sentence corpora. Establishing and using such a corpus for scenes could elucidate what kind of structure and rules are more than just superficially analogous to language rules, and for which rules a new and completely independent nomenclature will be necessary.

Despite many examples of the modulatory power of scene context on object processing, it still remains largely unclear what 'ingredients' of an object's context influence its processing. There is evidence that the mere summary statistics of a scene can modify object perception [43,44]. Moreover, knowledge regarding spatial positioning of not only target, but also distractor objects acquired through a lifetime of seeing objects in specific configurations speeds object search and perception. Kaiser *et al.* [45•] showed that it was easier to find targets in displays in which the distracters can be grouped (e.g. mirror above sink), and Gronau and Shachar [46] showed that contextual consistency facilitated LTM of perceptual detail in images shown for only a glimpse of an eye.

Using fMRI and MEG decoding, Brandmann and Peelen [47•] demonstrated that expectations derived from scene information and processed in scene-selective cortex, feedback to shape object representations in visual cortex implying functional interactions in space and time between scene-processing and object-processing pathways. With regard to the nature and time course of the interplay of semantic and syntactic predictions, Stein and Peelen [48] found independent influences of category-based and spatial attention on object detection (at least for familiar object categories). In a recent MEG study, Battistoni *et al.* [49] showed that naturalistic category search is initially guided via spatially global category processing, which then guides spatial attention to the location of the target. Thus, different types of predictions (e.g. regarding a scene's meaning and structure) seem to interact rapidly and on various levels of visual processing to enhance the efficiency of real-world attention, perception, and memory. How are such predictions or 'priors' stored in long-term memory as part of our scene grammar?

Scene grammar: a hierarchy of anchored predictions?

We neither believe, nor claim that the grammar that governs scene processing is the same as the grammar that allows us to understand and produce language. For instance, language is unique to humans and some linguistic operations (e.g. active/passive constructions) or linguistic categories (e.g. nouns, verbs) cannot easily be translated into scene components. However, there are also some commonalities between scene and language processing.

A somewhat provocative hypothesis proposes that the properties which scene and language processing have in common strongly depend on a hierarchically structured set of rules, that is, a grammar (see Ref. [50] for a review). In language, a so-called parser constitutes a program for analyzing a string of words (i.e. a sentence) and assigning it a structure in accordance with the rules of syntax, or more broadly speaking the rules of grammar (for a review see Ref. [51]). Similarly, the efficient processing of scenes requires a cognitive program that parses a scene into meaningful elements by applying the rules of a visual grammar. Both linguistic and scene grammar allow for the understanding of an infinite number of new sentences or new scenes, respectively. Recently, it has been shown that also goal-directed actions can be sequenced in small units, which are organized according to a hierarchical plan, resembling the hierarchical organization of language [52]. Moreover, there has been some converging evidence that the prefrontal cortex (PFC), specifically, BA44, may function as the essential region for hierarchical processing across the domains [53]. A real investigation of whether domain-general computations are shared between scene and language processing requires that the basic architecture of the grammar — which parses scenes and activates predictions — must first be identified. The next important milestones, therefore, lie in deciphering the nature of such predictions and how they are generated. This should include a particular emphasis on identifying the key components of such a scene grammar that allow for efficient guidance of both attention and perception.

Objects that make up a scene are not created equal. For instance, some objects tend to be more important for scene categorization, that is, those are more 'diagnostic', than others [54•,55–57]. When obscuring even large portions of a scene, 'the presence of a single diagnostic object is sufficient to rescue recognition' [58]. Similarly, we argue that so-called 'anchor objects' play a key role in scene perception, object search, as well as object identification. Anchors tend to be prominent objects that are diagnostic for a scene, for example the shower in the bathroom or the stove in the kitchen. However, their most important feature is that other objects within that scene have defined spatial relations with regard to these anchors. That is, while diagnostic objects tell us *what scene* we are in, anchors can tell us *where objects* are. For example, the shampoo *in* the shower, the pot *on* the stove, the lamp *beside* the bed [59]. That is, anchors can predict the location of many other objects in the scene. Thus, there tend to be separate groups of objects clustering around different anchors in the same scene creating in themselves meaningful units or 'phrases', for example, separate shower versus toilet versus sink phrases within a single bathroom scene (see Figure 2). When you look for a toothbrush you will, therefore, probably quickly exclude the shower and toilet phrases from your search and focus your attention on the sink.

We have recently operationalized this initially vague concept of anchors through four determinants: 1) the frequency in which objects appear together, 2) the distance between objects, 3) the variance of the spatial location, and 4) clustering of objects within scenes. By applying such an algorithm to large labeled databases, for example, LabelMe [60], we were able to identify and manipulate anchor objects in 3D rendered scenes to test their behavioral effects on attention allocation. Eye movements showed that presence of anchors is crucial in guiding search in naturalistic scenes [61•].

One important part in delineating the hierarchical structure in scenes will be the probing of the directionality of these nested predictions within and across hierarchy levels (does the sink predict the toothpaste to the same degree that the toothpaste predicts the sink?) as well as a more detailed investigation of how object functions divide a scene into meaningful subregions [12]. Furthermore, do different anchor objects within a scene predict each other horizontally within the same level (does the toilet predict the shower) to the same degree as they predict the global scene (the bathroom) and other objects (the toilet paper) vertically across levels? And what are the contributions of visual similarity, size, and spatial distance to the strength of these predictions? The set of rules forming a scene's grammar and governing perception is vast. While linguistic grammars have been extensively studied going as far back as to ancient languages like Sanskrit, our understanding of a scene's grammar is — despite some pioneering work of the 70 s — quite rudimentary and still awaits more systematic investigations. We hope this review will aid in sparking more interest in pushing research along this frontier.

Conclusion

Bridging the gap between studies of cognition using highly abstract artificial tasks and simplified stimuli and the complex, diverse realities of the world will offer important insights. As such, we need a new type of ecological perspective [62], one that values well-controlled laboratory research but seeks to understand how we make sense of and interact with our actual environment. Understanding the efficiency of object search and perception in real-world scenes by determining, for example, the hierarchical structure of our predictions could open up new translational horizons and opportunities, for example, by developing more sophisticated computer algorithms or by providing a diagnostic marker for impairments of other rule-governed learning like dyslexia even before children start school.

Conflict of interest statement

Nothing declared.

Acknowledgement

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — project number 222641018 — SFB/TRR 135, subproject C7 to MLV.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest

1. DiCarlo JJ, Zoccolan D, Rust NC: **How does the brain solve visual object recognition?** *Neuron* 2012, **73**:415-434 http://dx.doi.org/10.1016/j.neuron.2012.01.010.

2. Greene MR, Oliva A: **Recognition of natural scenes from global properties: seeing the forest without representing the trees**. *Cogn Psychol* 2009, **58**:137-176 http://dx.doi.org/10.1016/j.cogpsych.2008.06.001.

3. Lowe MX, Rajsic J, Ferber S, Walther DB: **Discriminating scene categories from brain activity within 100 milliseconds**. *Cortex* 2018, **106**:275-287 http://dx.doi.org/10.1016/j.cortex.2018.06.006.

4. Oliva A: **Gist of the scene**. In *Neurobiology of Attention.* Edited by Itti L, Rees G, Tsotsos JK. 2005:251-256.

5. Potter MC, Staub A, O'Connor DH, Potter MC: **Pictorial and conceptual representation of glimpsed pictures**. *J Exp Psychol Hum Percept Perform* 2004, **30**:478-489 http://dx.doi.org/10.1037/0096-1523.30.3.478.

6. Itti L, Koch C: **Computational modelling of visual attention**. *Nat Rev Neurosci* 2001, **2**:194-203 http://dx.doi.org/10.1038/35058500.

7. Malcolm GL, Shomstein S: **Object-based attention in real-world scenes**. *J Exp Psychol Gen* 2015, **144**:257-263 http://dx.doi.org/10.1037/xge0000060.

8. Gatys LA, Kümmerer M, Wallis TSA, Bethge M: *Guiding Human Gaze with Convolutional Neural Networks*. . Retrieved from 2017 *http://arxiv.org/abs/1712.06492*.

9. • Adeli H, Vitu F, Zelinsky GJ: **A model of the superior colliculus predicts fixation locations during scene viewing and visual search.** *J Neurosci* 2016, **37**:1453-1467 http://dx.doi.org/10.1523/jneurosci.0825-16.2016. Retrieved from *http://www.jneurosci.org/content/early/2016/12/30/JNEUROSCI.0825-16.2016*.

This is a great example of vision and language joining forces to create a model that can explain signature eye movement findings in both fields of research.

10. • Tatler BW, Brockmole JR, Carpenter RHS: **LATEST: a model of saccadic decisions in space and time**. *Psychol Rev* 2017, **124**:267-300 http://dx.doi.org/10.1037/rev0000054.

The LATEST is not only a sequel to Carpenter's LATER model, but finally allows to predict not only where but also when people look in one unified model.

11. Henderson JM, Malcolm GL, Schandl C: **Searching in the dark: cognitive relevance drives attention in real-world scenes**. *Psychon Bull Rev* 2009, **16**:850-856 http://dx.doi.org/10.3758/PBR.16.5.850.

12. Castelhano MS, Witherspoon RL: **How you use it matters: object function guides attention during visual search in scenes**. *Psychol Sci* 2016, **27**:606-621 http://dx.doi.org/10.1177/0956797616629130.

13. Henderson JM: **Regarding scenes**. *Curr Dir Psychol Sci* 2007, **16**:219-222 http://dx.doi.org/10.1111/j.1467-8721.2007.00507.x.

14. Henderson JM, Hayes TR, Rehrig G, Ferreira F: **Meaning guides attention during real-world scene description**. *Sci Rep* 2018, **8**:13504 http://dx.doi.org/10.1038/s41598-018-31894-5.

15. Võ ML-H, Henderson JM: **The time course of initial scene processing for eye movement guidance in natural scene search**. *J Vis* 2010, **10**:14 http://dx.doi.org/10.1167/10.3.14.

16. Eckstein MP, Koehler K, Welbourne LE, Akbas E: **Humans, but not deep neural networks, often miss giant targets in scenes**. *Curr Biol* 2017, **27**:2827-2832.e3 http://dx.doi.org/10.1016/j.cub.2017.07.068.

17. Draschkow D, Võ ML-H: **Of "what" and "where" in a natural search task: active object handling supports object location memory beyond the object's identity**. *Atten Percept Psychophys* 2016, **78**:1574-1584 http://dx.doi.org/10.3758/s13414-016-1111-x.

18. Võ ML-H, Wolfe JM: **The role of memory for visual search in scenes**. *Ann N Y Acad Sci* 2015, **1339**:72-81 http://dx.doi.org/10.1111/nyas.12667.

19. • Biederman I, Mezzanotte RJ, Rabinowitz JC: **Scene perception: detecting and judging objects undergoing relational violations**. *Cogn Psychol* 1982, **14**:143-177.

In this paper, Biederman *et al.* systematically discussed and tested effects of various object-scene inconsistencies on object perception abilities while introducing the terms semantic versus syntactic violations.

20. Võ ML-H, Henderson JM: **Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception**. *J Vis* 2009, **9**:24.1-15 http://dx.doi.org/10.1167/9.3.24.

21. Adolph KE, Kretch KS: **Gibson's theory of perceptual learning**. *International Encyclopedia of the Social & Behavioral Sciences*. 2015:127-134 http://dx.doi.org/10.1016/B978-0-08-097086-8.23096-1.

22. Gibson EJ: *Learning to Perceive or Perceiving to Learn? In International Society for Ecological Psychology*. . Oxford, OH 1989.

23. Malcolm GL, Henderson JM: **Combining top-down processes to guide eye movements during real-world scene search**. *J Vis* 2010, **10**:1-11 http://dx.doi.org/10.1167/10.2.4.

24. Võ ML-H, Wolfe JM: **The interplay of episodic and semantic memory in guiding repeated search in scenes**. *Cognition* 2013, **126**:198-212 http://dx.doi.org/10.1016/j.cognition.2012.09.017.

25. Võ ML-H, Henderson JM: **Object-scene inconsistencies do not capture gaze: evidence from the flash-preview moving-window paradigm**. *Atten Percept Psychophys* 2011, **73**:1742-1753 http://dx.doi.org/10.3758/s13414-011-0150-6.

26. Wolfe JM, Võ ML-H, Evans KK, Greene MR: **Visual search in scenes involves selective and nonselective pathways**. *Trends Cogn Sci* 2011, **15**:77-84 http://dx.doi.org/10.1016/j.tics.2010.12.001.

27. Davenport JL, Potter MC: **Scene consistency in object and background perception**. *Psychol Sci* 2004, **15**:559-564 http://dx.doi.org/10.1111/j.0956-7976.2004.00719.x.

28. Munneke J, Brentari V, Peelen MV: **The influence of scene context on object recognition is independent of attentional focus**. *Front Psychol* 2013, **4**:552 http://dx.doi.org/10.3389/fpsyg.2013.00552.

29. Henderson JM, Weeks PA, Hollingworth A: **The effects of semantic consistency on eye movements during complex scene viewing**. *J Exp Psychol Hum Percept Perform* 1999, **25**:210-228 http://dx.doi.org/10.1037/0096-1523.25.1.210.

30. Cornelissen THW, Võ ML-H: **Stuck on semantics: processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior**. *Atten Percept Psychophys* 2016, **79**:154-168 http://dx.doi.org/10.3758/s13414-016-1203-7.

31. Greene MR, Fei-Fei L: **Visual categorization is automatic and obligatory: evidence from Stroop-like paradigm**. *J Vis* 2014, **14** http://dx.doi.org/10.1167/14.1.14.

32. Li FF, VanRullen R, Koch C, Perona P: **Rapid natural scene categorization in the near absence of attention**. *Proc Natl Acad Sci U S A* 2002, **99**:9596-9601 http://dx.doi.org/10.1073/pnas.092277599.

33. Cohen MA, Alvarez GA, Nakayama K: **Natural-scene perception requires attention**. *Psychol Sci* 2011, **22**:1165-1172 http://dx.doi.org/10.1177/0956797611419168.

34. Firestone C, Scholl BJ: **Cognition does not affect perception: evaluating the evidence for "top-down" effects**. *Behav Brain Sci* 2016, **39**:e229 http://dx.doi.org/10.1017/S0140525X15000965.

35. Ganis G, Kutas M: **An electrophysiological study of scene effects on object identification**. *Brain Res Cogn Brain Res* 2003, **16**:123-144. Retrieved from *http://www.ncbi.nlm.nih.gov/pubmed/12668221*.

36. Võ ML-H, Wolfe JM: **Differential electrophysiological signatures of semantic and syntactic scene processing**. *Psychol Sci* 2013, **24**:1816-1823 http://dx.doi.org/10.1177/0956797613476955.

37. Mudrik L, Lamy D, Deouell LY: **ERP evidence for context congruity effects during simultaneous object-scene processing**. *Neuropsychologia* 2010, **48**:507-517 http://dx.doi.org/10.1016/j.neuropsychologia.2009.10.011.

38. Mudrik L, Shalgi S, Lamy D, Deouell LY: **Synchronous contextual irregularities affect early scene processing: replication and extension**. *Neuropsychologia* 2014, **56**:447-458 http://dx.doi.org/10.1016/j.neuropsychologia.2014.02.020.

39. Truman A, Mudrik L: **Are incongruent objects harder to identify? The functional significance of the N300 component**. *Neuropsychologia* 2018, **117**:222-232 http://dx.doi.org/10.1016/j.neuropsychologia.2018.06.004.

40. Draschkow D, Heikel E, Vo ML-H, Fiebach CJ, Sassenhagen J: **No evidence from MVPA for different processes underlying the N300 and N400 incongruity effects in object-scene processing**. *Neuropsychologia* 2018, **120**:9-17 http://dx.doi.org/10.1016/j.neuropsychologia.2018.09.016.

41. • Cohn N, Jackendoff R, Holcomb PJ, Kuperberg GR: **The grammar of visual narrative: neural evidence for constituent structure in sequential image comprehension**. *Neuropsychologia* 2014, **64**:63-70 http://dx.doi.org/10.1016/j.neuropsychologia.2014.09.018.

Neil Cohn has been at the forefront of laying out a grammar for visual narratives, like Peanuts comics and has shown interesting similarities in EEG responses to violations of such a grammar with language processing.

42. Maffongelli L, Bartoli E, Sammler D, Kolsch S, Campus C, Olivier E, Fadiga L, D'Ausilio A: **Distinct brain signatures of content and structure violation during action observation**. *Neuropsychologia* 2015, **75**:30-39 http://dx.doi.org/10.1016/j.neuropsychologia.2015.05.020.

43. Brady TF, Shafer-Skelton A, Alvarez GA: **Global ensemble texture representations are critical to rapid scene perception**. *J Exp Psychol Hum Percept Perform* 2017, **43**:1160-1176 http://dx.doi.org/10.1037/xhp0000399.

44. Lauer T, Cornelissen THW, Draschkow D, Willenbockel V, Vo ML-H: **The role of scene summary statistics in object recognition**. *Sci Rep* 2018, **8**:14666 http://dx.doi.org/10.1038/s41598-018-32991-1.

45. • Kaiser D, Stein T, Peelen MV: **Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex**. *Proc Natl Acad Sci U S A* 2014, **111**:11217-11222 http://dx.doi.org/10.1073/pnas.1400559111.

This paper shows that it's easier to find targets in displays, in which the distracters can be grouped (e.g. mirror above sink). This might partly explain the efficiency of naturalistic search, in which distracters are arranged in regular configurations.

46. Gronau N, Shachar M: **Contextual consistency facilitates long-term memory of perceptual detail in barely seen images**. *J Exp Psychol Hum Percept Perform* 2015, **41**:1095-1111 http://dx.doi.org/10.1037/xhp0000071.

47. • Brandman T, Peelen MV: **Interaction between scene and object processing revealed by human fMRI and MEG decoding**. *J Neurosci* 2017, **37**:7700-7710 http://dx.doi.org/10.1523/JNEUROSCI.0582-17.2017.

Using a clever combination of fMRI and MEG Decoding, Talia Brandmann and Marius Peelen demonstrate that expectations derived from scene information and processed in scene-selective cortex, feed back to shape object representations in visual cortex.

48. Stein T, Peelen MV: **Object detection in natural scenes: independent effects of spatial and category-based attention**. *Atten Percept Psychophys* 2017, **79**:738-752 http://dx.doi.org/10.3758/s13414-017-1279-8.

49. Battistoni E, Kaiser D, Hickey C, Peelen MV: **Spatial attention follows category-based attention during naturalistic visual search: evidence from MEG decoding**. *bioRxiv* 2018 . Retrieved from *http://biorxiv.org/content/early/2018/08/14/390807.abstract*.

50. Fitch WT: **Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition**. *Phys Life Rev* 2014, **11**:329-364 http://dx.doi.org/10.1016/j.plrev.2014.04.005.

51. Everaert MBH, Huybregts MAC, Chomsky N, Berwick RC, Bolhuis JJ: **Structures, not strings: linguistics as part of the cognitive sciences**. *Trends Cogn Sci* 2015, **19**:729-743 http://dx.doi.org/10.1016/j.tics.2015.09.008.

52. Maffongelli L, Antognini K, Daum MM: **Syntactical regularities of action sequences in the infant brain: when structure matters**. *Dev Sci* 2018, **21**:e12682 http://dx.doi.org/10.1111/desc.12682.

53. Jeon H-A: **Hierarchical processing in the prefrontal cortex in a variety of cognitive domains**. *Front Syst Neurosci* 2014, **8**:223 http://dx.doi.org/10.3389/fnsys.2014.00223.

54. • Greene MR: **Statistics of high-level scene context**. *Front Psychol* 2013, **4**:777 http://dx.doi.org/10.3389/fpsyg.2013.00777.

This paper offers a rich and interesting variety of descriptive statistics regarding object-scene relationships.

55. Biederman I: **On the semantics of a glance at a scene perceptual organization**. In *Perceptual Organization.* Edited by Kubovy M, Pomerantz J. Hillsdale: Laurence Earlbaum Associates; 1981:213-253.

56. Biederman I, Mezzanotte RJ, Rabinowitz JC: **Scene perception: detecting and judging objects undergoing relational violations**. *Cogn Psychol* 1982, **14**:143-177 http://dx.doi.org/10.1016/0010-0285(82)90007-X.

57. Friedman A: **Framing pictures: the role of knowledge in automatized encoding and memory for gist**. *J Exp Psychol Gen* 1979, **108**:316-355. Retrieved from *http://www.ncbi.nlm.nih.gov/pubmed/528908*.

58. MacEvoy SP, Epstein RA: **Constructing scenes from objects in human occipitotemporal cortex**. *Nat Neurosci* 2011, **14**:1323-1329 http://dx.doi.org/10.1038/nn.2903.

59. Draschkow D, Võ ML-H: **Scene grammar shapes the way we interact with objects, strengthens memories, and speeds search**. *Sci Rep* 2017, **7**:16471 http://dx.doi.org/10.1038/s41598-017-16739-x.

60. Russell BC, Torralba A, Murphy KP, Freeman WT: **LabelMe: a database and web-based tool for image annotation**. *Int J Comput Vis* 2008, **77**:157-173.

61. • Boettcher SEP, Draschkow D, Dienhart E, Võ M-H: **Anchoring visual search in scenes: assessing the role of anchor objects on eye-movements during visual search**. *J Vis* 2018, **18** http://dx.doi.org/10.1167/18.13.11.

This is the first paper actually testing the behavioral importance of anchors during real-world search. It also includes a first attempt to operationalize anchors using object statistics computed from a labeled image database.

62. Gibson JJ: *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin; 1979.