Chapter 2
What Is Attention?

Matei Mancas and Vincent P. Ferrera

2.1 The Study of Attention: A Transversal Approach

For humans, attention is a self-evident mental phenomenon that is active during every single moment of awareness. It was studied first by philosophers, followed by experimental psychologists, cognitive psychologists, and cognitive neuroscientists. Lately, computer scientists have discovered the usefulness of attention for modeling behavior in humans and machines and for better learning to solve complex tasks. These areas of study emerged sequentially, but add one on top of the others like the layers of an "attention onion" (Fig. 2.1).

Due to the highly diverse applications of attention, a precise and general definition is not easy to find. Moreover, views on attention have evolved over time and research domains. This chapter is structured into two parts. In the first part, we briefly survey the long history of attention-related research from philosophy to cognitive psychology, to which were added cognitive neuroscience and computer science. The second part of the chapter covers different aspects of attention in an attempt to arrive at a working definition that can be applied across multiple disciplines.

M. Mancas (✉)
Numediart Institute, University of Mons, Mons, Belgium
e-mail: matei.mancas@umons.ac.be

V. P. Ferrera
Zuckerman Institute on Mind Brain and Behavior, Columbia University, New York, NY, USA

M. Mancas et al. (eds.), *From Human Attention to Computational Attention*,
https://doi.org/10.1007/978-3-031-84300-6_2

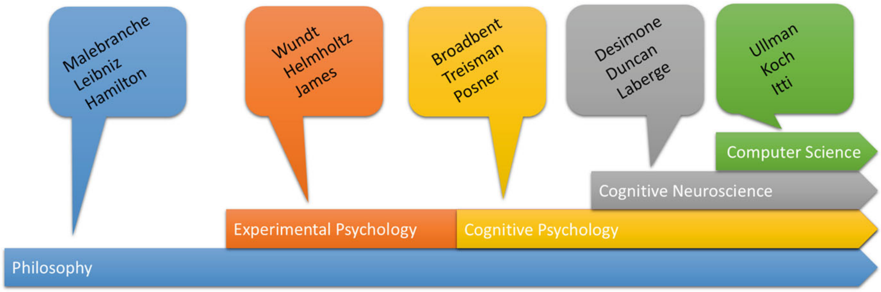

Fig. 2.1 History of attention and the attention onion: an accumulation of research domains

2.2 A Short History of Attention

Attention seems almost absent from any serious writing until the modern age. How did philosophers overlook such a key concept from the ancient times to the Enlightenment? Part of the answer is probably that attention is such a self-evident part of life that nobody really noticed it until experimental psychologists started to study sensory and perceptual processes quantitatively in the nineteenth century. Even if the concept of attention might be found in Greek philosophy [1], it is not the concept itself which is studied, but it is seen as a way to deal with weak senses and distractors to access the Truth. In the same way, meditation which exists in most of the ancient traditions uses attention as a tool to avoid awakened people to be distracted by futile details which prevent them from seeing the most important. We might then think that the idea of selective attention is very ancient and related to the philosopher/mystic/awakened who is able to master his selective attention to access Truth/Divine which is the most important and is the only worthy of focused attention. As a side effect he gets peace/ataraxia/apatheia (undivided cognitive load focused only on the most important which decreases the mental workload). While attention issues might lead to ignorance, they might also lead to madness (*mania*) in Plato which is an interesting link to attention-related pathologies.

2.2.1 Conceptual Findings: Attention in Philosophy

An early and important inquiry into human attention was that of Nicolas Malebranche, a French Oratorian priest who was also a philosopher and follower of Descartes. In his *De la Recherche de la Verité* (Concerning the Search after Truth) published in 1675, Malebranche focused on the role of attention in providing structure in scene understanding and the organization of thought. He also saw attention as the basis of free will, writing that "the occasional cause of the presence of ideas is attention . . . and it is easy to recognize that this is the principle of our freedom" [2]. Thus, from the very beginning attention was seen as linked to volition and consciousness.

In the eighteenth century, G. W. Leibniz developed the concept of "apperception" which refers to the assimilation of new and past experience into a current view of the world [3]. Leibniz' intuited an involuntary form of attention (known today as "bottom-up" or "stimulus-driven"), which is *needed for a perceived event to become conscious*. In Leibniz' view, attention is a reflexive and involuntary gate to consciousness.

In the nineteenth century, Sir W. Hamilton, a Scottish metaphysician, challenged the previous view on attention, which consisted in thinking that humans can only focus on a single stimulus at once. Hamilton noted that when people throw marbles, the locations of about seven of the marbles could be remembered [4]. This finding opened the way to the notion of "divided attention." The limited span of divided attention led about one century later to the famous paper of G.A. Miller, "The Magical Number Seven, Plus or Minus Two" in 1956 [5].

2.2.2 Attention in Experimental Psychology

After these early philosophical investigations, attention entered a scientific phase with the emergence of experimental psychology in the nineteenth century. By studying individual differences in the ability of trained astronomers to judge the transit of a celestial body through a telescope, W. Wundt introduced the study of consciousness and attention to the field of psychology [6]. He interpreted the observation error as the time needed to *switch one's attention voluntarily from one stimulus to another* and initiated a series of studies on the speed of mental processing, made possible by new methods of measuring simple and choice reaction times pioneered by F. Donders [7]. Here attention comes to be related to deliberative decision-making and not reflex alone.

In the second half of the nineteenth century, H. Von Helmholtz, in his "Treatise on Physiological Optics" [8] noted that despite the illusion of seeing our entire visual environment at the same spatial resolution, humans need to move their eyes around the whole visual field "because that is the only way we can see as distinctly as possible all the individual parts of the field in turn." Although his experimental work mainly involved the analysis of the eye movement scanpaths (overt attention), he also noted the existence of a covert attention, which is the ability to focus on different parts of a scene without moving the eyes. Von Helmholtz focused on the role of attention as an answer to the question of *"where" the objects of interest are*. Adding to the concepts of reflex attention and divided attention, the notion of parallel versus serial processing was born.

In 1890, W. James published his textbook *The Principles of Psychology* [9] and remarked again that attention is closely related to consciousness and structure. According to James, attention makes people perceive, conceive, distinguish, and remember. It also shortens reaction time. He linked attention to the notion of data compression and memory. He also developed a taxonomy of attention that distinguished between "passive" and "voluntary" attention. Contrary to Von Helmholtz, James was more focused on the fact that attention should answer the question of *"what" are the objects of interest*.

2.2.3 Attention in Cognitive Psychology

Between the very beginning of the twentieth century and 1949, the mainstream approach in psychology was behaviorism, which focused almost exclusively on the external causes of behavior. During this period, the study of mind was considered as barely scientific and few important advances were achieved in the field of attention. Despite this "hole" in the study of attention, important work was done on so-called interference effects. One of the most famous examples, the "Stroop Effect," was reported by J. R. Stroop [10], who showed that reaction times are considerably lengthened when a single stimulus affords two conflicting responses, for example reading a word such as "GREEN" (written in RED) or "RED" (written in GREEN) where the color of the world does not fit with the word itself and being asked to report the color of the ink in which the word was printed. This takes longer and is more error prone than reporting the ink color of "GREEN" (written in GREEN) or "RED" (written in RED), where the colors and words are congruent. Attention was invoked as a means to resolve the response conflict and explain the differences in performance between conflicting and congruent stimuli.

After the Second World War, a vastly more technological world emerged. Advances in information theory, statistical decision theory, and, perhaps most importantly, digital computing gave rise to the information age. Human performance in complex environments ranging from battlefields to factory floors became a central concern. The study of attention made a tremendous comeback. To the behaviorist view, which states that the organism's behavior is controlled by stimulus-response-outcome associations, cognitive psychology showed that behavior can be modulated by attention. The resurgence of attention began with the work of C. Cherry in 1953 on the "cocktail party" paradigm [11]. This approach models how people select the conversation that they are listening to and ignore the rest. This problem was called *"focused attention," as opposed to "divided attention."*

In the late 1950s, D. Broadbent [12] proposed a "bottleneck" model in which he described the selective properties of attention. His idea was that *attention acts like a filter (selector) of relevant information* based on basic features, such as color or orientation for images. If the incoming information matches the filter it can reach awareness (conscious state); otherwise it will be discarded. At that time, the study of attention seemed to become very coherent and was called "early selection." Nevertheless, after this short positive period, most of the findings summarized by Broadbent proved to be conflicting.

The first "attack" came from the alternative model of Deutsch and Deutsch [13] who used some properties of the cocktail-party paradigm to introduce a *"late* *selection" model*, where attentional selection is basically a matter of memory processing and response selection. The idea is that all information is acquired, but only that which fits semantic or memory-related objects is selected to reach awareness. This view is opposite to Broadbent who professed an early selection of the features before they reach any further processing.

New models were introduced like the attenuated filter model of A. Treisman [14] which is a softer version than Broadbent's bottleneck and which allows stimuli with a *response higher than a given threshold through the filter*, thus determining the focus of the selective attention.

Later, in 1980, Treisman and Gelade [15] proposed a new "feature integration" theory, where attention occurs in two distinct steps. First, *a preattentive parallel effortless step* analyzes objects and extracts features from those objects. In a second step, those *features are combined to obtain a hierarchy of focus* attention which pushes information towards awareness.

Despite its importance, the feature integration theory was also highly disputed. Other theories emerged as M. Posner [16] proposed a *spotlight supporting a spatial selection* approach, or D. Kahneman [17] and his theory of capacity supporting the idea of *mental effort*.

In the late 1980s, a plethora of theories on attention flourished and none of them was capable of accounting for all previous findings. According to H. Pashler [18], after several decades of research in cognitive psychology, more questions were raised than answers given. As a provocative rejoinder to the famous "Everyone knows what attention is" proposed by James a century before, Pashler declared that "No one knows what attention is."

2.2.4 The Need for New Approaches: After the Late 1980s "Crisis"

Attention deals with the allocation of cognitive resources to prioritize incoming information in order to bring them to a conscious state, update memory and the scene model, and influence behavior. Between consciousness, memory, and behavior, attention was revealed to be much more complex than initially expected and some people even questioned whether attention was a single concept or, rather, that there are several different forms of attention. The number of issues and the complexity of the nature of attention led to an interesting move in splitting attention studies from one single community into multiple different communities.

The cognitive neuroscience community has the goal of getting further into the theoretical and biological nature of attention using simple stimuli. The arrival of advanced tools such as functional imaging, EEG, MEG, or single-cell recordings in awake, behaving subjects allows them to make huge steps towards relating neural recordings with behavioral correlates of attention.

The segment of the computer science community working in the field of attention has a goal of making the concept work with real data such as images, videos, audio, or 3D models. From the late 1990s and the first computational models of visual attention, the cognitive neuroscience and computer science approaches have developed in parallel, one trying to get more insight on the biological brain and the other trying to get results which can predict eye movements and other behavior for real-life stimuli and environments. Even if the computational attention community led to some models very different from what is known to happen in the brain, the engineers' creativity is impressive and the results on real-life data begin to be significant and the applications endless.

2.2.5 Attention in Cognitive Neuroscience

Cognitive neuroscience arrived with its own set of tools and methods. If some of them were already used in cognitive psychology (e.g., EEG, eye-tracking devices), others are new tools providing new insights on brain and behavior:

- Psychophysiological methods: scalp recording of EEG (electroencephalography—measures the large-scale electric activity of the neurons) and MEG (magnetoencephalography—measures the magnetic fields produced by electrical currents in the brain) which are complementary in terms of sensitivity on different brain areas of interest.
- Neuroimaging methods: functional MRI and PET scan images which both measure the areas in the brain which have intense activity given a task that the subject executes (visual, audio, etc.). Magnetic resonance spectroscopy can provide information about specific neurotransmitters.
- Electrophysiological methods: single-cell recordings, which measure the electrophysiological responses of a single neuron or population of neurons using a microelectrode system. While this system is much more precise, it is also more invasive.
- Other methods: TMS and TDCS (transcranial magnetic stimulation and transcranial direct current stimulation, which can be used to stimulate a region of the brain and to measure the activity of specific brain circuits in humans) and multi-electrode technology which allows the study of the activity of many neurons simultaneously showing how different neuron populations interact and collaborate.

Using those techniques two main families of theories have been constructed.

The first and most well-known model is the biased competition model of Desimone and Duncan in [19]. The central idea is that at any given moment, there is more information in the environment than can be processed. Relevant information always competes with irrelevant information to influence behavior. Attention biases this competition, increasing the influence of behaviorally relevant information and decreasing the influence of irrelevant information. Desimone and Duncan explicitly suggest a physiologically plausible neural mechanism that mediates this competition for the visual system. A receptive field of the neuron is a window to the outside world. The neuron reacts only to stimuli in this window and is insensitive to stimulation in other areas. The authors assume that the competition between stimuli takes place if more than one stimulus is present in the same receptive field. This approach is very interesting as each neuron can be seen as a filter by itself and the neuron receptive field can vary from small and precise (like in the primary visual cortex V1) to large enough to focus on entire objects (like higher visual areas in the temporal and parietal lobes). This basic idea suggests different domains of attention (location-based, feature-based, object-based, attentional bottleneck) in a very natural and elegant way. Moreover, a link is achieved with memory based on the notion of attentional templates in working memory which enhance neuronal responses depending on previously acquired data. This idea is embodied in the selective tuning model of Tsotsos in 1995 [20].

A second family of models was developed by Laberge in late 1990s [21]. It is a structural model based on neuropsychological findings and data from neuroimaging studies. Laberge conjectures that at least three brain regions are concurrently involved in the control of attention: frontal areas, especially the prefrontal cortex; thalamic nuclei, especially the pulvinar and posterior sites; and the posterior parietal cortex in the intraparietal sulcus. Laberge proposes that these regions are necessary for attention and all these regions presumably give rise to attentional control together. While cognitive neuroscience brought a lot of new methods and information to cognitive psychology, attention is still far from being fully understood and a lot of work is ongoing in the field.

2.2.6 Attention in Computer Science

While cognitive neuroscience focuses on researching the biological nature of attention, a different angle arose in the 1980s with the improvements in computational power. Building on the Feature Integration Theory of Treisman and Gelade [15], C. Koch and S. Ullman [22] proposed that the different visual features that contribute to attentive selection of a stimulus (color, orientation, movement, etc.) are combined into one single topographic map, called the “saliency map.” The saliency map integrates the normalized information from the individual feature maps into one global measure. Bottom-up saliency is determined by how different a stimulus is from its surroundings at several scales. The saliency map provides the probability for each region in the visual field to be attended. This saliency map concept is close to that of the “master map” postulated in the feature integration theory by Treisman and Gelade.

The first computational implementation of Koch and Ullman architecture was achieved by Laurent Itti in his seminal work [23]. This very first computational implementation of an attention system takes as an input any image and outputs a saliency map of this image and also the winner-take-all-based mechanism simulating the eye fixations during scene analysis. Following Itti, hundreds of models were developed first for images, then for videos, and some for audio or even 3D data very recently.

From the initial biologically inspired models a number of models based on mathematics, statistics, or information theory arrived on the "saliency market," making better and better predictions about human attention. These models are all based on features extracted from the signal (most of the time low-level features but not always), such as luminance, color, orientation, texture, motion, objects' relative position or even simply neighborhoods, or patches from the signal. Once those features are extracted, all the existing methods are essentially based on the same principle: looking for "contrasted, rare, surprising, novel, worthy-to-learn, less compressible, or information maximizing" areas. All those terms suggest searching for some unusual features in a given context. This context can be local (typically center-surround spatial or temporal contrasts) and global (whole image or very long temporal history), or it can be a model of normality (the image average, the image frequency content). Very recently learning is more and more involved in computing saliency: first, it was mainly about adjusting model coefficients given a precise task; now complex classifiers like deep neural networks are beginning to be used to both extract the features from the signal and train the most salient features based on ground truth obtained with eye-tracking or mouse-tracking data.

The arrival of deep neural networks has provided tools to enhance attention models especially on their top-down dimension [24]. But very deep neural networks learn a kind of implicit attention [25] which helps them focus on part of a data stream and provide better results.

In addition to that, attention modules were coded into neural networks, especially the transformer architectures [26]. Those modules help to make better decisions and are also good in multimodal approaches where parts of different signals are important at one moment to take a final decision. The arrival of attention mechanisms with the learning architectures themselves is an important step and a chapter in this book will be focused on this topic.

2.3 So . . . What Is Attention?

The transdisciplinary nature of attention naturally leads to a lot of different definitions. Attention deals with the allocation of cognitive resources to prioritize incoming information in order to bring them to a conscious state, update a scene model, update memory, and influence behavior. But several attention mechanisms were highlighted especially considering Cherry's cocktail party phenomenon. A dichotomy appeared between divided attention and selective attention. From there, clinical observations led to a model of attention divided into five different "kinds." One can also talk about different kinds of attention that rely on overt gaze or not, or that use only image features versus memory and emotions. While its purpose seems to be the relation between the outer world and inner consciousness, memory and emotions, the clinical manifestation of attention and attention deficits tends to show that there might be several varieties of attention.

2.3.1 Overt vs. Covert: The Eye

Overt versus covert attention is a distinction that was noted at the very beginning of psychological studies on attention. Overt attention is manifested by changes in posture that prepare sensory receptors for expected input. Eye movements, head movements, external ear (pinna) movements, changes in pupil size, and so forth are all examples of overt attention. Covert attention does not induce eye movements or other postural changes: it is the ability to catch (and thus be able to bring to consciousness) regions of a scene which are not fixated by the eyes. Eye movements are necessitated by the non-uniform distribution of receptive cells (cones and rods) on the retina. The cones which provide a high resolution and color are mainly concentrated in the middle of the retina in a region called "fovea." This means that in order to acquire a high spatial resolution image the eye must gaze towards the area or object of interest to align it on the fovea. This constraint led to mainly three types of eye movements:

1. Fixations: the gaze stays a minimal time period on approximately the same spatial area. The eye gaze is never still. Even when gazing at a specific location, small movements such as tremor, slow drifts, and microsaccades can be detected. The microsaccades are very small movements of the eye during area fixations.
2. Saccades: the eyes make a ballistic movement between two fixations. They disengage from one fixation and they are very rapidly shifted to the second fixation within about 1/20th of a second. Between the two fixations, no visual data is acquired. This is called saccadic suppression.
3. Smooth pursuit: a smooth pursuit is like fixation on a moving object. The eye will follow a moving object to maintain it in the fovea (central part of the retina). During smooth pursuits, more rapid small corrections, i.e., "corrective saccades," can be done to correct position errors.

Modeling overt attention attempts to predict human fixation locations and the dynamical path of the eye (called the eye "scanpath").

2.3.2 Serial vs. Parallel: The Cognitive Load

While focused, sustained, and selective attention deal with the serial processing of information, alternating and divided attention deal with parallel processing. These distinctions suggest that attention can deal with information both serially and in parallel. While there is a limit to the number of tasks that are processed in parallel during divided attention (around five tasks), in the case of preattentive processing, massively parallel computation can be done. Some notions such as scene “gist” [27] seem to be very fast and encompass the entire visual field to get a first and very rough idea about the context of the environment. The five kinds of attention follow a hierarchy based on the degree of focus, thus the cognitive load which is needed to achieve the attentive task. This approach is sometimes called the clinical model of attention:

1. Focused attention: respond to specific stimuli (focus on a precise task).
2. Sustained attention: maintain a consistent response during longer continuous activity (stay attentive a long period of time and follow the same topic).
3. Selective attention: selectively maintain the cognitive resource on specific stimuli (focus only on a given object while ignoring distractors).
4. Alternating attention: switch between multiple tasks (stop reading to watch something).
5. Divided attention: deal simultaneously with multiple tasks (talking while driving).

2.3.3 Bottom-Up vs. Top-Down: Memory and Actions

Another fundamental property of attention needs to be taken into account: attention is a mix of two components referred to as bottom-up (or exogenous) and top-down (or endogenous). The bottom-up component is reflexive and is driven by the acquired signal. Attention is attracted by the novelty of some features in a given context (spatial-local, a contrasted region; spatial-global, a red dot while all the others are blue; temporal, a change in direction or speed of movement). Its main purpose is to alert in case of unexpected or rare situations and it is tightly related to survival. This first component of attention is the one which is the best modeled in computer science as the signal features are objective cues which can be easily extracted in a computational way.

The second component of attention (top-down) deals with individual subjective feelings. It is related to rewards, memory, emotions, and individual goals. This component of attention is less easy to model by computers as it is more subjective and it requires information about internal states, goals, a priori knowledge, or emotions. Top-down attention can be itself divided into two sub-components:

1. Goal/action-related: Depending on an individual current goal, certain features or locations are inhibited and others receive more weight. The same individual with the same prior knowledge responds differently to the same stimuli when the task in hand is different. This component is also called “volitional.”
2. Memory/emotion-related: This process is related to experience and prior knowledge (and the emotions related to them). In this category one can find the scene context (experience from previously viewed scenes with similar spatial layouts or similar motion behavior) or object recognition (you see your grandmother first in the middle of other unknown people). This component of attention is more

"automatic"; it does not impose a large cognitive load and it can come along with volitional attention. The other way around, volitional top-down attention cannot inhibit the memory-related attention which will still work whether a goal is present or not. More generally, bottom-up attention cannot be inhibited if there is a strong and unusual signal acquired. If someone searches for his keys (volitional top-down), he will not take care about a car passing by. But if he hears a strange sound (bottom-up) and then recognizes a lion (memory-related top-down attention), he will stop searching for the keys and run away. Volitional top-down attention is able to inhibit the other components of attention only if they are not very intense.

2.4 Attention vs. Attentions: A Summary

The study of attention is an accumulation of disciplines ranging from philosophy to computer science and passing through psychology and neuroscience. Those disciplines often study different aspects or views of attention, which leads to a situation where a single and precise definition of attention is simply not feasible.

To sum up the different approaches attention is about:

- Eye/neck/head movements and outside world information acquisition: the attentional "embodiment" leads to parallel and serial attention (overt versus covert attention).
- Allocation of cognitive resources to important incoming information: the attentional "filtering" is the first step towards data structuring (degree of focus and clinical model of attention).
- Mutual influence on memory and emotions: passing of important information to a conscious state and getting feedback from memory and emotions (bottom-up and memory-related top-down attention).
- Behavior update: reacting to novel situations but also managing the goals and actions (bottom-up and volitional top-down attention).
- Global human well-being: focus and distractors (selective versus divided attention).

Attention plays a crucial role, partly conscious and partly unconscious, from signal acquisition to action planning going through the main cognitive steps. Or maybe there are simply several attentions and not only one. At this point in time, this question still has no final answer.

References

1. Guo, Z. (2022). *Attention: A preliminary history of the philosophical topic from Homer to Aristotle*. (Master's thesis).

2. Greenberg, S. (2008). Things that undermine each other: Occasionalism, freedom, and attention in Malebranche. In D. Garber & S. Nadler (Eds.), *Oxford studies in early modern philosophy* (Vol. 4, pp. 113–140). Oxford University Press.
3. Runes, D. D. (Ed.). (2001). *The dictionary of philosophy*. Citadel Press.
4. Hamilton, W. (1859). *Lectures on metaphysics and logic* (Vol. 1). Gould and Lincoln.
5. Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81.
6. Wundt, W. M. (1904). *Principles of physiological psychology* (Vol. 1). Sonnenschein.
7. Goldstein, E. (2014). *Cognitive psychology: Connecting mind, research and everyday experience*. Cengage Learning.
8. von Helmholtz, H. (2005). *Treatise on physiological optics* (Vol. 3). Courier Corporation.
9. James, W. (1913). *The principles of psychology* (Vol. II). Henry Holt and Co.
10. Jensen, A. R., & Rohwer, W. D. (1966). The Stroop color-word test: A review. *Acta Psychologica, 25*, 36–93.
11. Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America, 25*(5), 975–979.
12. Broadbent, D. E. (1957). A mechanical model for human attention and immediate memory. *Psychological Review, 64*(3), 205.
13. Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70*(1), 80.
14. Treisman, A. M. (1964). Selective attention in man. *British Medical Bulletin, 20*, 12.
15. Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136.
16. Posner, M. I. (1995). *Attention in cognitive neuroscience: an overview*. University of Oregon.
17. Friedenberg, J., & Silverman, G. (2011). *Cognitive science: An introduction to the study of mind*. Sage.
18. Pashler, H. E., & Sutherland, S. (1998). *The psychology of attention* (Vol. 15). MIT press.
19. Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*(1), 193–222.
20. Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (n.d.). Modeling visual attention via selective tuning. *Artificial Intelligence, 78*(1), 507–545.
21. Laberge. (1999). Networks of attention. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences*. MIT Press, 2004.
22. Koch, C., & Ullman, S. (1987). Shifts in selective visual attention: Towards the underlying neural circuitry. In *Matters of intelligence* (pp. 115–141). Springer.
23. Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(11), 1254–1259.
24. Kummerer, M., Wallis, T. S., Gatys, L. A., & Bethge, M. (2017). *Understanding low-and high-level contributions to fixation prediction* (pp. 4789–4798). In Proceedings of the IEEE International Conference on Computer Vision.
25. Mnih, V., Heess, N., & Graves, A. (2014). Recurrent models of visual attention. *Advances in Neural Information Processing Systems, 27*.
26. Vaswani, A. (2017). Attention is all you need. *arXiv preprint arXiv:1706.03762*.
27. Torralba, A., et al. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113*(4), 766.