Phenomenology and Architecture: Examining Embodied Experience and Graphic Representations of the Built Environment

Jennifer A. E. Shields

> …the task of architecture extends beyond its material, functional, and measurable properties – and even beyond aesthetics – into the mental and existential sphere of life. (Juhani Pallasmaa, Architect, 2015)

We can describe a work of architecture by its objective characteristics – descriptors that are independent of an occupant or viewer. These characteristics include location, physical context, style, age, dimensions, proportions, and materiality. An objective description is quickly exhausted as dependent qualities come into view. These include the function or use, and the meaning or symbolism found in the design. These dependent qualities are notable not only because of their subjectivity but because of their variability. Meaning or symbolism intended by the architect at the time of a building's construction may be lost or transformed over time. Architecture may take on new meaning to future occupants. The Pantheon in Rome was valued as a temple to the Roman gods by the ancient Romans, as a Christian church by medieval Romans, and as a mausoleum beginning in the Renaissance. Architects and engineers today admire it as an early concrete structure, which remains the largest unreinforced concrete dome in the world (Fig. 1).

The independent and dependent characteristics of a work of architecture indicate the distinction between the objective world and the world as an individual being experiences and understands it. Edmund Husserl described these two worlds in *Thing and Space: Lectures of 1907*. Here he says that there is a spatial and temporal world around us, but that we are the "centers of reference" for this world. "The environing objects [Objekten], with their properties, changes, and relations, are what they are for themselves, but they have a position relative to us, initially a spatio-temporal position and then also a 'spiritual' [= cultural] one" (cited in Woodruff Smith, 2013, 219). Similarly, Maurice Merleau-Ponty describes this duality in his course notes from the Collège de France in the 1950s, building on biologist

J. A. E. Shields (✉)
California Polytechnic State University, San Luis Obispo, CA, USA
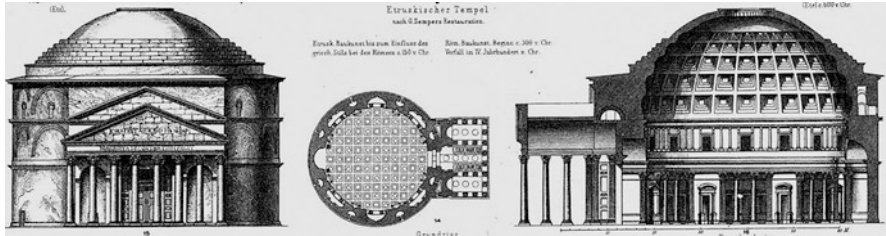e-mail: jeshield@calpoly.edu

Fig. 1 Pantheon elevation, plan, and cross-section drawings, 1885, Meyers Konversationslexikon, Vierte Auflage, author unknown. (Public domain)

Jacob von Uexküll's 1930s concept of the *Umwelt*. "The *Umwelt* marks the difference between the world such as it exists in itself, and the world of a living being. It is an intermediary reality between the world as it exists for an absolute observer and a purely subjective domain" (Merleau-Ponty & Séglard, 2003, 167).

The intersection between the objective and subjective defined here becomes even more dynamic with the inclusion of graphic representations. Graphic representations of the built environment include drawings, renderings, and photographs, among other media. Architects sometimes use *graphic representation* and *image* interchangeably, but we understand that there are different meanings and connotations by discipline. French philosopher Gaston Bachelard describes multiple meanings of the term *image*: "…the word *image*, in the works of psychologists, is surrounded with confusion: we see images, we reproduce images, we retain images in our memory" (Bachelard et al., 1994, xxxiv). These internal and external images are part of a feedback loop between our perceptions and representations. As author Italo Calvino declared in 1985, well before the ubiquity of social media, "We live in an unending rainfall of images" (Calvino, 1988, 57). Our exposure to external images has increased exponentially. This rainfall of images includes many representations of the built environment. Some are published by design professionals, but the vast majority of images are produced and proliferated through Instagram and other social media platforms. Our experience of the world is mediated through the framed, cropped, and filtered photos we see on a daily basis. These images act as a mediator between the individual and the built environment, further shaping our *Umwelt*.

Von Uexküll uses the metaphor of intersecting soap bubbles to visualize the relationship between each organism's *Umwelt*. He describes a meadow in order to understand the myriad worlds of humans and animals.

> …we must first blow, in fancy, a soap bubble around each creature to represent its own world, filled with the perceptions which it alone knows. When we ourselves then step into one of these bubbles, the familiar meadow is transformed. Many of its colorful features disappear, others no longer belong together but appear in new relationships. A new world comes into being. Through the bubble we see the world of the burrowing worm, of the butterfly, or of the field mouse; the world as it appears to the animals themselves, not as it appears to us. This we may call the *phenomenal world or the self-world of the animal*…We thus unlock the gates that lead to other realms, for all that a subject perceives becomes his *perceptual world* and all that he does, his *effector world*. Perceptual and effector worlds together form a closed unit, the *Umwelt*. (Von Uexküll, 1992, 319–320)

For humans, the delimitation of our phenomenal world based on *perceptual* and *effector* tools is not a complete description – images are a filter between the individual and the world, altering both what we perceive and what we do. As architects, we create images for different purposes and audiences, including envisioning how people might experience and use a space. We produce drawings to work out a design, but also to convince clients and future users of the viability of our proposal. As authors not just of the architecture but also of some of its representations, it is critical to understand the implications in representing architecture in different graphic modes. René Magritte called attention to the subjectivity of graphic representations in 1948 in his painting entitled *The Treachery of Images*. The painting shows a pipe, beneath which it says "Ceci n'est pas une pipe" ("This is not a pipe"). With this graphic and text, Magritte is pointing out that what he painted is not an actual pipe, but only his interpretation or representation of a pipe. All modes of graphic representation in architecture are abstractions of physical space and form, and so they are subject to interpretation.

This chapter will draw on theories and studies from a range of disciplines to describe the relationship between architecture and embodied experience, and how graphic representations mediate our experience of the built environment. Part 1 will discuss characteristics of our embodied experience and graphic representations of the built environment, while Part 2 will present processes of visual perception and how our visual perception of both the built environment and its representations have been empirically studied (Fig. 2).

1 Part 1

1.1 The Built Environment: Embodied Experience and Graphic Representation

Our environment, be it built or natural, offers more stimuli than we can process or even perceive. Our selective attention to the vast array of stimuli defines our *Umwelt*. Our individual experience of the built environment is bound by our *effector* capacities and our *perceptual* capacities, to use von Uexküll's subsets of the *Umwelt*. What we do is shaped by affordances, and what we perceive is shaped by multi-sensory perception and memory. Part 1 will break down these capacities as well as the roles and modes of graphic representation.

1.2 Affordances

Affordance theory, developed by psychologist James J. Gibson, describes the relationship between the world and the occupant in a similar way to von Uexküll's *effector* world of the *Umwelt*. Gibson says: "The affordances of the environment are what
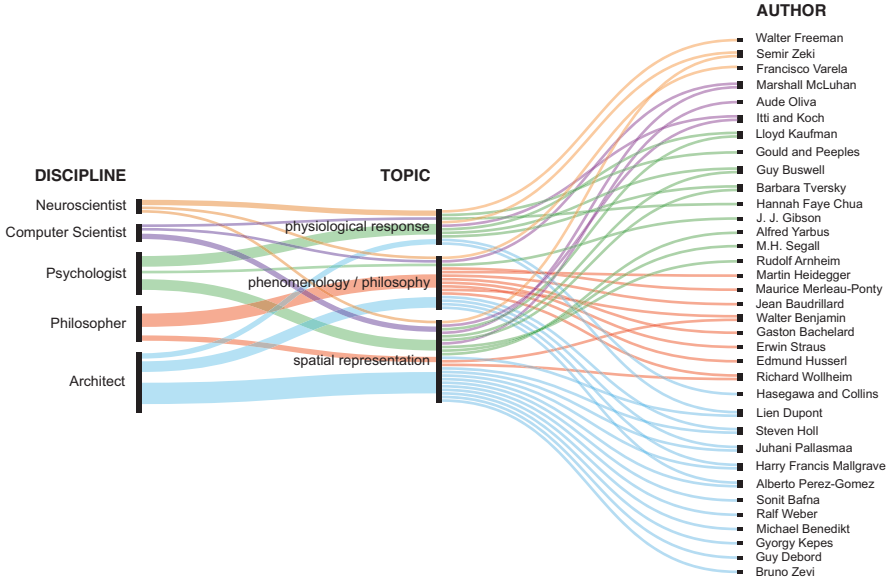
Fig. 2 Interdisciplinary research diagram. (By author)

it *offers* the animal, what it *provides* or *furnishes*, either for good or ill. The verb *to afford* is found in the dictionary, the noun *affordance* is not. I have made it up" (1979, 127). From this definition, we see that while the built environment has objective characteristics, the perception and use of it changes depending on our age, culture, abilities, and needs. The affordances that we identify in our environment extend in part from what is sometimes called our "sixth sense" – proprioception. Proprioception is the term for our understanding of our body in space. "Proprioception, or kinesthesia, is the sense that lets us perceive the location, movement, and action of parts of the body…It combines with other senses to locate external objects relative to the body…" (Taylor, 2009, 1143). Our ability to judge the position of our body in space can only develop through practice – we have to move in the world and interact with it for our bodies to understand and make sense of it. This is also true for our bodies and tools for movement, such as bicycles, skateboards, and wheelchairs. In 1911, neurologists Henry Head and Gordon Morgan Holmes developed the concept of the *body schema*, which is related to proprioception. Body schema refers to: "the body's relations to immediately surrounding space. The body schema includes the brain and sensory processes that register the posture of one's body in space. The body schema is plastic, amenable to constant revision, extends beyond the envelope of the skin, and has important implications for tool use" (Robinson, 2015, 138). The body schema incorporates *effector* tools, like bicycles and wheelchairs, which become extensions of our bodies. Due to the variety of ways we can interact with the built environment, at different speeds, with different footprints, the designer has to anticipate many types of possible actions in the designed space.

Our own understanding of our body in space and how it relates to the built environment is shaped by proprioception, but also by cultural norms and expectations. For example, a chair affords sitting for an adult, while it affords climbing, standing, and jumping for a child. These expectations include how we engage with the physical structures of the built environment, as well as how we interact with each other in these spaces. Proxemics, a concept developed by anthropologist Edward Hall in the 1960s, describes the study of the spatial separation between people, depending on the type of interaction. "If the spatial experience is different by virtue of different patterning of the senses and selective attention and inattention to specific aspects of the environment, it would follow what crowds one people does not necessarily crowd another" (Hall et al., 1968, 84). The comfort level for various types of interactions varies culturally, and can also be transformed by external forces such as the coronavirus and expectations of social distancing, which impact the distance at which we feel comfortable standing from one another. Our *effector* world – what we do – is impacted by the affordances we perceive.

1.3   Multi-sensory Perception

Bridging affordances and multi-sensory perception – the *effector* and *perceptual* worlds – is philosopher Walter Benjamin's theory (first published in 1935) of a twofold appropriation of architecture. Benjamin (1968) states,

> Architecture has always represented the prototype of a work of art the reception of which is consummated by a collectivity in a state of distraction… Buildings are appropriated in a twofold manner: by use and by perception – or rather, by touch and sight. Such appropriation cannot be understood in terms of the attentive concentration of a tourist before a famous building. On the tactile side there is no counterpart to contemplation on the optical side. Tactile appropriation is accomplished not so much by attention as by habit. As regards architecture, habit determines to a large extent even optical reception. The latter, too, occurs much less through rapt attention than by noticing the object in incidental fashion. This mode of appropriation, developed with reference to architecture, in certain circumstances acquires canonical value. For the tasks which face the human apparatus of perception at the turning points of history cannot be solved by optical means, that is, by contemplation, alone. They are mastered gradually by habit, under the guidance of tactile appropriation. (339–340)

As an occupant of a work of architecture, particularly a building we inhabit regularly, our perception of it is dominated by our use of and interaction with the space. The tactile sense supersedes the visual. The overwhelming amount of sensory data in our environment – new or familiar – means that we must pay attention to some things while ignoring others, known as selective attention. Novel stimuli attract our attention. Thus a familiar place causes habituation, the diminishing of a physiological response due to repeated stimuli. Through this process, we no longer 'see' the building: we instead develop a muscle memory for our physical interactions with it.

Juhani Pallasmaa, a Finnish architect who advocates for multi-sensory experience in architectural design, published a seminal book on the topic in 1996 called

The Eyes of the Skin: Architecture and the Senses. In it, he says: "Every touching experience of architecture is multi-sensory; qualities of space, matter and scale are measured equally by the eye, ear, nose, skin, tongue, skeleton and muscle. Architecture strengthens the existential experience, one's sense of being in the world…" (Pallasmaa, 2012, 45). Pallasmaa proposes that the sense of alienation and isolation often felt in contemporary architecture and urban spaces is a result of a lack of engagement of the senses. The dominance of the visual (driven in part by our image-saturated society) positions us as observers rather than participants in the world. Pallasmaa holds as exemplary the work of Swiss architect Peter Zumthor, who approaches his designs through a phenomenological lens. Zumthor describes the importance of material selection and detailing to provide a multi-sensory experience. "The sense that I try to instill into materials is beyond all rules of composition, and their tangibility, smell, and acoustic qualities are merely elements of the language that we are obliged to use. Sense emerges when I succeed in bringing out the specific meanings of certain materials in my buildings, meanings that can only be perceived in just this way in this one building" (Zumthor, 2010, 10). In Zumthor's thermal baths in Vals, Switzerland, the dim lighting in the interior augments other sensory experiences, including the thermal and tactile qualities of the water and stone, and the subtle light and shadow migrating through the space (Fig. 3).

Fig. 3 Peter Zumthor, Therme Vals, Switzerland, 1996. (Watercolor of the floral bath by Thomas di Santo, 2004)

FLORAL BATH M/ FLOWER PETALS + FRAGRANCE IN WATER   THERME HOTEL/VALS
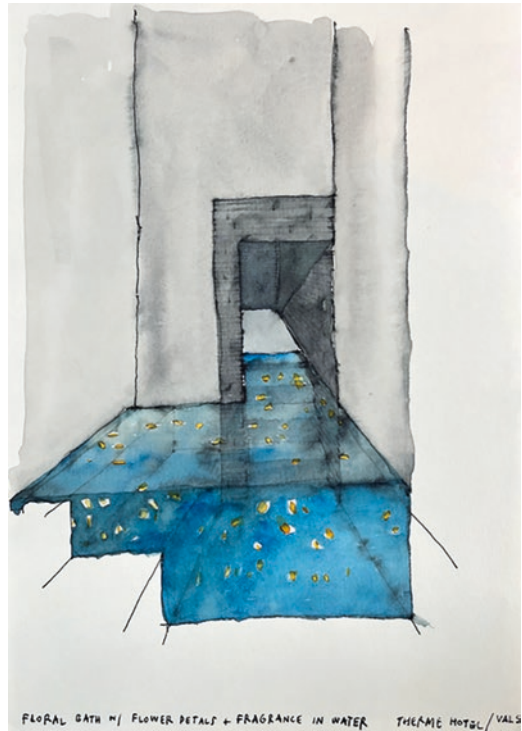
1.4   Memory

Sensory perception cannot be separated from memory. In the Merriam-Webster dictionary, perception is defined as: "physical sensation interpreted in the light of experience" (2011). The intertwining of direct perceptual experience with memory is highlighted in Zumthor's written and built work. While Zumthor values the direct sensory impact of architecture, he also describes the importance of memory on his design process. In his reflections, he recalls specific childhood sensory experiences, like the feel of a smooth metal door handle. He states: "Memories like these contain the deepest architectural experience that I know. They are the reservoirs of the architectural atmospheres and images that I explore in my work as an architect" (Zumthor, 2010, 8). Zumthor's work, while inspired by his personal experiences and memories of architecture, is intended to provide space for the occupant's own memories to surface. While Zumthor describes architecture as potentially offering space for reflection and reminiscence, Bachelard describes the inverse in his treatise on architecture and phenomenology, *The Poetics of Space.* Architecture can provoke the recollection of memories, but architectural space can also help us to form new memories. He says: "Memories are motionless, and the more securely they are fixed in space, the sounder they are" (Bachelard et al., 1994, 9). There is a long history of tying memory to place, going back to the memory palace (method of loci) of the Greeks and Romans. This mnemonic device works by first visualizing a familiar place (your home, for example), and then picturing the items or topics you want to remember in a series of rooms. The great orator Cicero used this technique to memorize speeches.

While individual memories color our personal perceptions of architectural space, collective memory also plays a role. In Victor Hugo's novel *The Hunchback of Notre-Dame*, first published in 1831, he presents the idea of architecture as a repository for a society's collective memory. "In fact, from the origin of things up to the fifteenth century of the Christian era inclusive [when the printing press was invented], architecture was the great book of mankind, the principal expression of man at his different stages of development, whether as strength or as intelligence" (Hugo et al., 1999, 193). Hugo describes the invention of the printing press as a turning point in history, transforming the way cultures record and transmit their knowledge and values. Before this point, "…the human race never had an important thought which it did not write down in stone" (Hugo et al., 1999, 199). As described in the introduction to this chapter, however, the meaning communicated by a work of architecture can change over time. The transformation of meaning over time applies to both the work of architecture itself, as well as its images. Bachelard describes the overlay of individual and collective memory in our understanding of images. "Great images have both a history and a prehistory; they are always a blend of memory and legend, with the result that we never experience an image directly. Indeed, every great image has an unfathomable oneiric depth to which the personal past adds special color" (Bachelard et al., 1994, 33). Architectural images include mental images produced through an embodied experience of a place or in reading written imagery, as well as external images including drawings, renderings, and photographs (which can then produce mental images). In the next section, these external images – graphic representations – are described.

1.5   Graphic Representation

Our individual embodied experience of the built environment is shaped by affordances, multi-sensory perceptions, and memory, as described in the previous sections. These describe a direct relationship between our bodies and the environment. External architectural images that we come into contact with create an additional mediation between ourselves and our world, interceding and transforming our *Umwelt*.

Architects utilize a variety of modes of graphic representation in their design process and to communicate their ideas to clients and future occupants. (Architects also use three-dimensional representations like physical models, but this chapter focuses on two-dimensional representations of three-dimensional space, as those are the most ubiquitous and far reaching.) Additionally, print and social media deliver numerous images of the built environment by authors ranging from developers to real estate agents to vacationers. These authors have differing agendas, and thus produce media with differing messages. Calvino's statement that "We live in an unending rainfall of images" continues: "The most powerful media transform the world into images and multiply it by means of the phantasmagoric play of mirrors. These are images stripped of the inner inevitability that ought to mark every image as form and as meaning, as a claim on the attention and as a source of possible meanings" (Calvino, 1988, 57). Our oblique perception of these images, generally not given our focused attention or thoughtful evaluation, impacts our experiences and expectations of the built environment nonetheless. Our perceptual and effector capacities as they affect our experience of the embodied environment were described in the previous sections; here we will consider how we experience two-dimensional images. The very act of flattening three-dimensional space into a two-dimensional image requires abstraction and interpretation by the author, thus opening it up to multiple interpretations. There are two broad conceptions of architectural images that must be unpacked: the twofold perception of images as described by Richard Wollheim, and the "provocative" versus "instructional" drawing as defined by Sonit Bafna.

We all understand architectural drawings as referential, representing something besides themselves – architectural space and form. Philosopher Richard Wollheim describes the perception of drawings as twofold (1973). We see the drawing as an object, but also as a representation of something else. Wollheim describes the viewer's interaction with an image as "seeing-in," which involves an awareness of the marks on the surface, the drawing as an object, and an awareness of some absent object – the item depicted.

Wollheim's observations speak to drawings in general, but there are many types of drawings within the practice of architecture. Architectural theorist Sonit Bafna talks about two types of drawings. The "drawing-as-provocation" is imaginative, a tool to explore possible forms and spaces. The "drawing-as-instruction" is notational, serving to communicate how a building is to be constructed (2008). The imaginative drawing might lead to the notational drawing, and then the actual
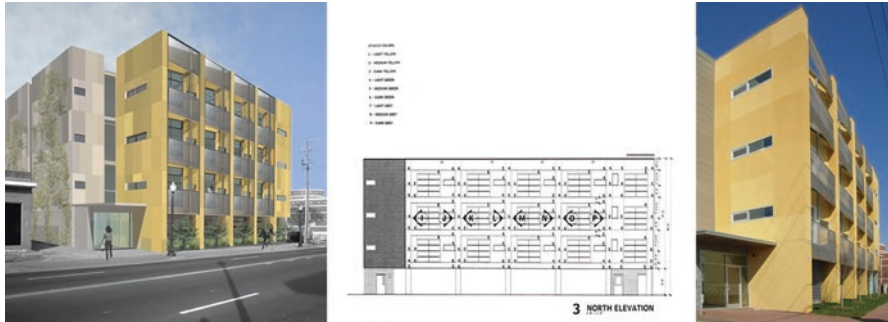
Fig. 4 Central 27 rendering, elevation, and photograph – the rendering on the left shows the stucco façade as having panels of different color shades – a result of the rendering software. The client liked this look, unintended by the architect, and so we produced a 'color-by-number' elevation drawing for the contractor to create this effect in the finished building. (Images by author)

building. Or we might stop with the imaginative drawing, either because a project isn't built, or because the image was intended for a rhetorical purpose. In practice, architectural drawings fall on a spectrum with Bafna's two types of drawings existing as the poles. Architects often develop their designs as follows: *imaginative sketches > hard-lined drawings > renderings > notational (construction) drawings* in order to produce a building.[1]

The imaginative drawing might seek to illustrate an anticipated atmosphere or experience. It is intended to explore one or more of the following: form, space, materiality, scale, light, and use, among other characteristics. The imaginative drawing can serve the client and intended users, but it is also a design tool for the architect. In contrast, a notational drawing is intended for a different audience. A notational drawing tends towards geometrical rather than experiential accuracy, providing instructions for building using discipline-specific conventions. Contractors rely on detailed, dimensioned, and annotated drawings to assemble the materials and systems called for in an architectural design. These drawings are most commonly orthographic views (plan, section, elevation), which are two-dimensional abstractions for the purpose of measuring and describing spatial and material relationships and methods of construction, rather than to evoke a space experientially. Because we do not perceive the world in orthographic, these drawings tend to require training for literacy in comprehension. Although these drawings are not produced to evoke the atmosphere or experience anticipated in a work of architecture, they are necessary to construct the building that will ultimately provide an embodied experience (Fig. 4).

---

[1] A sketch refers to a quick free-hand drawing. A hard-lined drawing refers to a measured, geometrically accurate drawing produced using a pencil and ruler or a computer-aided drafting program. A rendering is a computer-generated or hand-drawn perspective view of a building or space that is intended to show what a space will look like, including light and shadow, materials, and an understanding of human scale.

Now that we've looked at how the built environment is experienced and graphically represented, Part 2 will look more specifically at our visual perception of the world and its images, as empirically studied.

2 Part 2

2.1 Visual Perception of the Built Environment

The difference between the world as experienced in the flesh and the world as experienced through images describes a variation on Merleau-Ponty's definition of *Umwelt*. Each situation balances objective data with subjective understanding. Part 2 of this chapter will address the visual perception of the built environment and its representations in two-dimensional images. While the engagement of other senses contributes to our perception of the built environment, images lie in the realm of the visual, and so visual perception provides an appropriate mode of comparison. Part 2 will cover the basics of visual perception of three-dimensional space, the visual perception of two-dimensional representations of space, eye-tracking as a tool for quantitatively measuring our visual perception, and empirical studies investigating these topics, including the author's own studies.

In the study of visual perception, it is understood that "Both the physical world and the perceptual world have structure" and can be defined geometrically (Hershenson, 1999, 2). While the physical world can be described by Euclidean geometry (Hershenson, 1999), the geometry of the perceptual world, which is viewer-dependent, changes based on the location and orientation of the viewer. In the perceptual world, object attributes such as size, shape, movement, direction, and position vary (Hershenson, 1999). The world we inhabit is three-dimensional – yet visual information about our world is interpreted by the brain through two-dimensional images, projected onto our retinas. The single image resulting from synthesizing two different monocular images is called binocular fusion. The disparity between the two retinal images is what gives us cues about depth perception, or three-dimensional relationships between things. Thus objects that are closer appear larger, while objects in the distance appear smaller (Fig. 5).

With more stimuli in the environment than we are capable of processing, our perceptual system relies on two levels of processing to acquire the information we need. "*Top-down and bottom-up processing* refers to the integration of information from one's own cognitive system (top-down) and from the world (bottom-up) to facilitate perception" (Carlson, 2010, 1011). *Top-down processing* relies on our prior knowledge and expectations, influencing our perception. Perceiving the environment with a specific task in mind is an example of this. *Bottom-up processing* refers to information we receive from direct sensory input. Aude Oliva, a professor at MIT, bridges human perception/cognition, computer vision, and cognitive neuroscience in her studies of visual perception. Her research investigates the processes
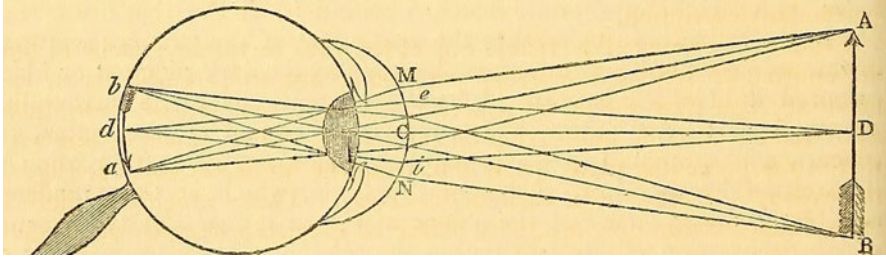
Fig. 5 diagram of visual perception by Joseph Chitty, in *A practical treatise on medical jurisprudence*, 1834. (Public domain)

by which we view a scene or place. In a first glance at a scene, "the visual system forms a spatial representation of the outside world that is rich enough to grasp the meaning of the scene, recognizing a few objects and other salient information in the image, to facilitate object detection and the deployment of attention" (Oliva, 2005, 251). She calls this the *gist* of a scene. Additionally, she describes the role of memory in this assessment:

> "…people rely on their previous experience and knowledge of the world to rapidly process the vast amount of detail in a real world scene. One's current view of a scene is automatically incorporated into a "scene schema," which includes stored memories of similar places which have been viewed in the past, as well as expectations about what is likely to be seen next. Although we aren't aware of it, viewing a scene is an active process, in which images are combined with memory and experience to create an internal reconstruction of the visual world." (Oliva, 2010, 1113)

In every perceptual experience, both levels of processing are at work. Top-down processing can be tied to von Uexküll's effector world, while bottom-up processing relates to the perceptual world. We employ top-down processing to a greater extent when we are in a familiar place, and we rely more on bottom-up processing when we are in an unfamiliar setting.

2.2 Visual Perception of 3D Space in 2D Representations

Top-down and bottom-up processing are also deployed as we look at two-dimensional representations of three-dimensional space. While architects and designers have graphic conventions for representation that are intended to communicate with others in related fields (like orthographic drawings, as mentioned earlier), we commonly use perspectival drawings and renderings to communicate with clients, future users, and the general public. Perspectival drawings are intended to approximate how a three-dimensional space would be visually perceived by the human eye – the perceptual world as opposed to the physical world as defined by Hershenson. These drawings place the viewer into the imagined space. Rudolf Arnheim, in his book *Art and Visual Perception: A Psychology of the Creative Eye*,

describes how we understand depth in a two-dimensional image. Perspective representation acknowledges a viewer: the distortion of the built environment occurs because there is a viewer and direction of view. "This explicit acknowledgment of the viewer is at the same time a violent imposition upon the world represented in the picture. The perspective distortions are not caused by forces inherent in the represented world itself. They are the visual expression of the fact that this world is being sighted" (Arnheim, 1956, 294). Here, the measurable geometry of the world is transformed, in order to approximate a visual experience.

He goes on to say: "Although the rules of central perspective produce pictures that closely resemble the mechanical projections yielded by the lenses of eyes and cameras, there are significant differences. Even in this more realistic mode of spatial representation, the rule prevails that no feature of the visual image will be deformed unless the task of representing depth requires it" (Arnheim, 1956, 285–286). We look for geometric simplicity to reconcile two- and three-dimensions. Arnheim defines the law of simplicity as our perceptual desire to find the simplest structure. In the diagram below, we could see the left figure as an irregular quadrilateral shape, but it makes more sense (it is formally simpler) as a square in perspective (Arnheim, 1956) (Fig. 6).

The interpretation of a two-dimensional line drawing intended to represent a view of three-dimensional space adheres to Arnheim's principles. However, many designers use renderings – perspectival drawings with light and shadow, color, and materiality – to more closely approximate an embodied experience. Little research has been done on our visual perception of architectural space in renderings. Renderings offer less room for interpretation and imagination by the viewer than perspectival line drawings because more attributes have been assigned within the rendered image. Yet even in seemingly realistic renderings, the correlation between the perception of the attributes of the rendering and the attributes of the built work is unknown.

Although we recognize drawings and renderings as an artist's or architect's depiction of space, with photography, there tends to be an assumption that it is an
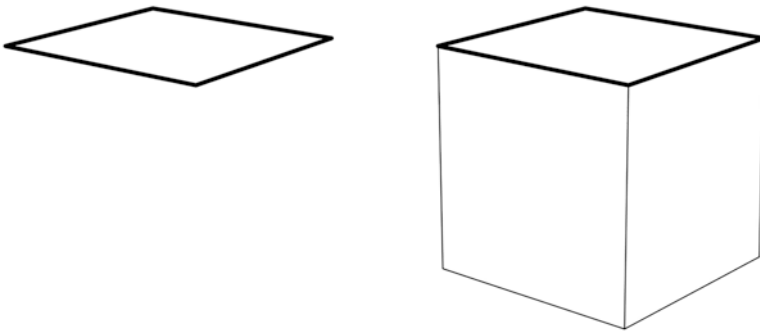

Fig. 6 Interpretation of a form in two and three dimensions, based on Arnheim's law of simplicity. (Diagram by author)

unbiased presentation of the space. Yet understanding the content and spatial implications of a photograph is a learned ability, and the photograph is, like a drawing, an abstraction of the architectural space. Studies comparing different cultures have found no difference in perceptual organization; however, these studies have shown that photographs are interpreted differently by different cultural groups (Weber, 1995). For example, African aborigines – who had never seen a photograph before – could not recognize anything in photographs they were shown, which depicted spaces familiar to them. Two-dimensional representations of three-dimensional space require interpretation which "relies on acquired visual conventions that may be as arbitrary as linguistic conventions" (Weber, 1995, 62). Photographs are authored and abstracted depictions of architectural space in the same manner as drawings. The photographer chooses the lens (which may or may not be similar to the lens of the human eye, and so may show more or less of the space), the direction of view, and what to include in or exclude from the frame. In some visual perception studies, a photograph serves as a stand-in for the real space, but the bias of the photograph is acknowledged. In order to better understand our visual perception of two-dimensional images, researchers often measure visual activity using eye tracking devices.

2.3   Visual Attention and Eye Tracking in Art and Architecture

Eye tracking is a valuable tool to quantitatively and spatially document a person's visual experience. Eye tracking data includes fixations and saccades, based on movement of the fovea, the area of the retina where eyesight is sharpest. The fovea represents less than 2 degrees of the visual field, so the eye must move, or *foveate*, in order to take in detailed information (for example, in complex areas of the visual field). When the eye stops to take in information, this is called a *fixation*. The rapid movement of the eye from one fixation to another is called a *saccade*. During saccades, the eye is effectively blind: visual information obtained by the eye occurs during fixations, thus making fixations the most valuable data collected through eye tracking (Holmqvist et al., 2015).

As early as the 1930s, psychologists were interested in how the eye moves and fixates when focused on a work of art, both with and without a given task. Guy Buswell, a psychologist who invented the first non-intrusive eye tracker, found that the eyes do not follow edges, but tend to scan and focus on central concave areas (1935). More recent studies confirm that fixations are likely to occur in concave and enclosed areas of a figure, rather than in what is perceived as negative space (Weber et al., 2002). Alfred Yarbus, another pioneer in eye tracking research, found that eye movements vary when looking at an image depending on the task the observer was asked to complete (1973). His most famous study asked participants to look at a painting called *The Unexpected Visitor* seven times, first without a prompt and then six more times based on prompts. Questions included details of the painting like remembering the position of people and objects in the room, as well as narrative

questions like estimating how long the visitor had been away from the family. The selective attention process caused participants to attend to certain aspects of the image while ignoring others. When asked to remember the position of people and objects in the room, participants viewed more of the painting with less attention to any one element, while when they were asked to estimate how long the visitor had been away from the family, participants predominantly looked at the faces of the people. Elina Pihko's 2011 eye tracking study compared how laypeople and experts look at paintings, determining that in a perspectival painting, the laypeople tend to fixate on the center of the space while the experts view more of the painting. These studies demonstrate the impact of top-down processing on the visual perception of images.

While the visual perception of artworks has been investigated, there has been little study on the role of eye movement in the perception of three-dimensional architectural space. One of the few studies on this topic was conducted by Weber et al. (2002) in which they collected eye tracking data as participants were asked to look at either scaled three-dimensional models or photographs of models of architectural spaces. The research focused on comparing different arrangements of objects within a space. The results of this study show that without a given task (dominated by bottom-up processing), the eye is drawn to visual centers and distinct objects rather than tracing contours. "Elements indicating spatial depth, such as vistas, receive special attention… [and] vertically and horizontally oriented objects are explored less than obliquely oriented shapes" (Weber et al., 2002, 57). This confirms the 'law of simplicity' proposed by Arnheim (1977), as we need time to process the meaning of angled lines in an image – do they indicate an irregular two-dimensional object, or a perspectival three-dimensional object? The results also confirm Buswell's conclusion that fixations are likely to occur in concave and enclosed areas of a figure, rather than in what is perceived as negative space (Weber et al., 2002). The study found that fixations did not vary significantly when viewing the three-dimensional model compared with a photograph of the model, with the exception of the foreground, which attracted greater attention in the physical model. This study shows that, when looking at representations of three-dimensional space, we attend to the features of the image that convey depth rather than to features of the image, like contours, that convey the two-dimensional composition of the image.

As mentioned earlier, the built environment provides substantially more sensory stimuli than we can attend to. In primates, visual information is carried along the optic nerve at a rate of approximately 10^8 bits per second. This, however, is too much for the brain to process into consciousness, so it has adapted by selecting certain parts of the image to process preferentially. This allows the brain to register different areas of interest in a serial manner and shift from one to another depending on the importance that the brain allots. *Saliency mapping* is a technique first conceived by neuroscientists Laurent Itti, Christof Koch, and Ernst Niebur at the California Institute of Technology in 1998. It is a method of computationally analyzing unique features in any given photograph and processing them to highlight the anticipated foci of attention. This visual attention system was created to simulate "the neuronal architecture of the early primate visual system" (Itti et al., 1998).

Saliency maps are created through learning algorithms that consider color, intensity and orientation. These saliency maps were originally created from a bottom-up perspective, and were subsequently modified by Torralba to take into account top-down processing (Torralba, 2005). A study by Torralba et al. (2006) considered the role of a given task in viewing an image. This study proposed visual attentional guidance through an experimental search task. Results of their study suggest that context plays an important role in object detection and observation. For example, if you are asked to look for a person in a scene of a busy urban street, you will likely look at street level, assuming a person would be there and not floating in mid-air. "…the scene priors constitute an effective shortcut for object detection as it provides priors for the object's presence/absence before scanning the image" (Torralba, 2005, 591). When studying visual perception, bottom-up and top-down processing cannot be easily isolated.

Lien Dupont et al. (2016) studied saliency maps in the context of landscape architecture. Landscape architects often conduct a *visual impact assessment* when proposing changes to a landscape. For example, this is required to minimize visual impact for changes along Route 1 on California's coast. With accurate computer-generated saliency mapping, landscape architects could determine whether their proposed design would significantly impact the visual perception and attention to a given landscape. In this study, computer-generated saliency maps were compared to human focus maps – obtained by collecting eye tracking data as participants viewed images – to test the accuracy of predicting the human viewing pattern. Seventy-four landscape photographs were shown, ranging from rural to urban scenes, for 10 s each. An eye tracking device recorded the results in the form of a focus map, which was then compared to the saliency map for each photograph. A relatively high correlation was found, demonstrating that saliency maps can be a reliable prediction of human's observation patterns. However, the correlation between the saliency map and focus map was found to be greater in rural landscapes, showing that human viewing behavior can be more easily predicted in these settings, as opposed to urban landscapes.

2.4 Eye Tracking to Compare Modes of Architectural Representation

The range of physiological responses evoked by both embodied experiences and the perception of architectural images demands a deeper investigation into their associations. Architects design with various modes of representation, and others outside the design professions represent the built environment through drawings and photography. As the audience varies, so does the way we represent things. What can the perception of images of architecture and cities tell us about the actual experience – or potential experience – of the places depicted? The long-term goal of the author's own empirical research attempts to relate physiological responses – like what we

focus on and what emotions we experience – in real spaces with the responses evoked when looking at architectural representations. The significance of this is that if architects can anticipate how people will experience a space based on how they respond to images of it, we could design better schools, hospitals, etc. (There will never be a perfect correlation, of course.) The following studies sit at the intersection of Architecture, Cognitive Science, and Phenomenology, where: "Phenomenology can also enrich our understanding of empirical results by embedding them in a coherent theoretical framework…In this mode, the phenomenologist is a kind of higher-level meta-theorist, drawing both on phenomenological and non-phenomenological sources in putting together an account of some kind of experiential process or pattern" (Yoshimi, 2016, 297).

In a pair of pilot studies conducted with architecture students in North Carolina, China, and California, we collected eye tracking data to test our preliminary hypothesis: that visual attention varies with representational mode (Shields et al., 2016). We compared eye tracking data for two different modes of representation – a perspectival line drawing and a matching photograph of the same space. We chose a space that did not have an obvious function associated with it and that had some spatial complexity. In order to analyze the eye tracking data, we identified seven Areas of Interest (AOIs) in the visual scene. Both studies found differences in how participants viewed the drawing compared with the photo. We interpreted the difference in visual attention as a distinction between spatial complexity and graphic complexity: our attention is drawn to more graphically complex areas (more lines) in a drawing but more spatially complex areas in a photo. This prompts the question: do we also look for spatial complexity in an embodied experience?

In a larger-scale study of the visual perception of architectural spaces, we were interested in both the mode of representation (comparing drawing to photograph) and the audience (comparing architecture student to preschool student). Through the collection of eye tracking data, our aim was to test our preliminary hypotheses: that salient features differ by representation mode, and by participant group. We used two different modes of representation, like the previous studies – a perspectival line drawing and a matching photograph of the Salk Institute in La Jolla, CA, by the architect Louis Kahn. In order to compare salient features, we set up Areas of Interest (AOIs), and collected eye movement data for each AOI. The participants were not given any instructions other than to look at the image, in order to privilege bottom-up processing (Fig. 7).

The results showed that architecture students' visual attention differed between the drawing and the photograph, while the attention of the preschool students did not. The results suggest that the trained eye (that of the architecture student) will focus on the architecture in an image, but to a greater degree in a drawing. Additionally, the eye movements of the architecture students support the architect's intent for the Salk Institute – that the architecture frame the sky and horizon, drawing the eye towards the ocean beyond. This is an interesting result given that the participants were not there in person but were looking at a photograph, in which the ocean is barely visible.
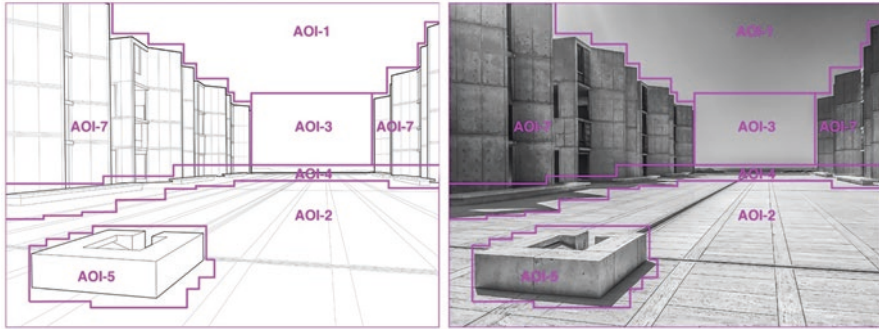
Fig. 7 Salk Institute drawing (L) and photo (R) with AOIs. (Images by author)

We also wanted to investigate whether a computer-generated saliency map could predict the areas of a photograph where most attention would be drawn. At first glance we could see that the saliency map predicts the ground as attracting attention, while the focus map from the participants shows the horizon and sky to be of greater interest. We would anticipate that a quantitative comparison would show the saliency map to differ from the focus map. Using Dupont el al.'s (2016) method for comparing a saliency map to a focus map, we arrived at a correlation coefficient of 0.3564, which indicates that while the two maps are somewhat similar, there is nonetheless a weak positive correlation between the two. From this, we can gather that the saliency map is only mildly reliable at predicting areas of interest, and cannot replace a focus map. This confirms Dupont et al.'s conclusion that the saliency map was more accurate for predicting attention in a rural landscape than an urban one.[2] The results suggest that the visuospatial literacy of the participant – the ability for the participant to perceive three-dimensional space in a two-dimensional image, rather than only two-dimensional characteristics – plays a role in what aspects of the photo are attended to. It could be possible for Artificial Intelligence to more closely approximate a human viewing pattern, with enough human eye tracking data to support machine learning (Fig. 8).

3   Conclusion

Through these experiments, we've found that both the mode of representation and the audience affect how architectural spaces are visually perceived, even without a task. There are additional types of physiological data we could collect – heart rate,

---

[2]The larger research project of which this study is a part is a cross-disciplinary investigation involving researchers in computer science, cognitive science, and architecture. This particular study was completed in consultation with Dr. Jeffrey Sklar, Statistics, and Dr. Laura Cacciamani, Psychology, with Research Assistants Wood Cheng, Gabriella Ojalvo, Melina van Oers.
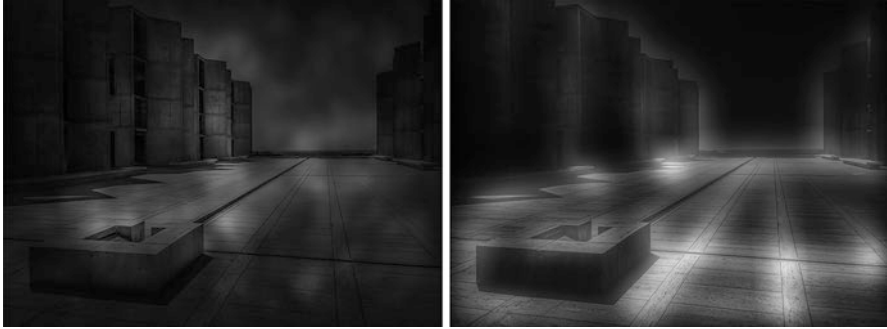
Fig. 8 Salk Institute focus map (L) and saliency map (R). (Images by author)

galvanic skin response, EEG – and many modes of representation to investigate, like renderings and virtual reality. Additionally, we could collect phenomenological data from the participants, asking them what they notice in an image. According to Francisco J. Varela in "Neurophenomenology: A Methodological Remedy for the Hard Problem," "One of the originalities of the phenomenological attitude is that it does not seek to oppose the subjective to the objective, but to move beyond the split into their fundamental correlation" (1996, 339). Varela's statement suggests that the subjective and objective can be correlated. Varela's conclusions in this essay have been criticized by Tim Bayne, but Bayne does propose that participant responses can "guide the analysis" of the quantitative data (2004).

An additional challenge is the measurement of physiological responses in an embodied experience. Photography, perhaps, more closely approximates the real space, but eliminates peripheral visual data and other sensory information and does not address the role of time in our experience of architecture. Since a static photograph is not a sufficient analog for the embodied experience, it would be valuable to collect physiological data from participants as they inhabit a work of architecture.

The empirical studies described here reinforce von Uexküll's claim (and the claims of Husserl, Merleau-Ponty, J.J. Gibson, and others) that we experience the world through the lenses of what we are able to perceive and what we are able to do in that world. In visual perception alone, our *Umwelt* is crafted by bottom-up (external stimuli) and top-down (internally motivated) processes, allowing us to 'see' some aspects of the world and be blind to others. Additionally, graphic representations of the world shape what we do and do not notice.

Calvino asks, "Will the power of evoking images of things that are *not there* continue to develop in a human race increasingly inundated by a flood of prefabricated images?" (1988, 91). While architects may not be able to stem the flood of images of the built environment proliferating across social media, we do have authorship over the drawings, renderings, and photographs we produce and share. The ostensible authority we have as designers of the built environment gives us a responsibility to critically consider our modes of representation and our audiences, and how one impacts the other. Perhaps Bachelard's suggestion can guide us. "Only

phenomenology – that is to say, consideration of the *onset of the image* in an individual consciousness – can help us to restore the subjectivity of image and to measure their fullness, their strength and their transsubjectivity" (Bachelard et al., 1994, xix).

While these philosophers note the individual and idiosyncratic nature of perception, the empirical research does bear out commonalities amongst test subjects. A valuable next step would be to expand both the objective measurement of physiological responses and the collection of phenomenological responses to embodied experience and graphic representations. Further research can employ both objective and subjective data in seeking to anticipate how people will respond to works of architecture. If we could anticipate how spaces will affect the occupant's phenomenological experience by how they respond to images of it, we could design a better built environment.

References

Arnheim, R. (1956). *Art and visual perception: A psychology of the creative eye*. Faber and Faber.

Arnheim, R. (1977). *The dynamics of architectural form*. University of California Press.

Bachelard, G., et al. (1994). *The poetics of space* (1994 edition). Beacon Press.

Bafna, S. (2008). How architectural drawings work – And what that implies for the role of representation in architecture. *The Journal of Architecture, 13*(5), 535–564.

Bayne, T. (2004). Closing the gap? Some questions for neurophenomenology. *Phenomenology and the Cognitive Sciences, 3*(4), 349–364.

Benjamin, W. (1968). The work of art in the age of mechanical reproduction. In *Illuminations*. Schocken Books.

Buswell, G. (1935). *How people look at pictures: A study of the psychology of perception in art*. University of Chicago Press.

Calvino, I. (1988). *Six memos for the next millennium*. Harvard University Press.

Carlson, L. (2010). Top-down and bottom-up processing. In B. Goldstein (Ed.), *Encyclopedia of perception* (pp. 1012–1014). SAGE. https://doi.org/10.4135/9781412972000.n324

Chitty, J. (1834). *A practical treatise on medical jurisprudence: With so much of anatomy, physiology, pathology, and the practice of medicine and surgery, as are essential to be known by members of parliament, lawyers, coroners, magistrates, officers in the army and navy*. H. Butterworth.

Dupont, L., et al. (2016). Comparing saliency maps and eye-tracking focus maps: The potential use in visual impact assessment based on landscape photographs. *Landscape and Urban Planning, 148*, 17–26.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

Hall, E., et al. (1968). Proxemics [and comments and replies]. *Current Anthropology, 9*(2/3), 83–108.

Hershenson, M. (1999). *Visual space perception: A primer*. MIT Press.

Holmqvist, K., et al. (2015). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Hugo, V., et al. (1999). *Notre-Dame de Paris*. Oxford University Press.

Itti, L., et al. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(11), 1254–1259.

Merleau-Ponty, M., & Séglard, D. (2003). *The nature: Course notes from the Collège de France*. Northwestern University Press.

Oliva, A. (2005). Gist of the scene. In L. Itti et al. (Eds.), *Neurobiology of attention* (pp. 251–256). Elsevier/Academic.

Oliva, A. (2010). Visual scene perception. In B. Goldstein (Ed.), *Encyclopedia of perception* (pp. 1111–1116). SAGE.

Pallasmaa, J. (2012). *The eyes of the skin: Architecture and the senses*. Wiley-Academy/Wiley.

Pallasmaa, J. (2015). Body, mind, and imagination: The mental essence of architecture. In S. Robinson & J. Pallasmaa (Eds.), *Mind in architecture: Neuroscience, embodiment, and the future of design*. The MIT Press.

perception. (2011). *In Merriam-Webster Dictionary*. https://www.merriam-webster.com/diction-ary/perception. Accessed 8 May 2020.

Pihko, E., et al. (2011). Experiencing art: The influence of expertise and painting abstraction level. *Frontiers in Human Neuroscience, 5*, 94.

Robinson, S. (2015). Nested bodies. In S. Robinson & J. Pallasmaa (Eds.), *Mind in architecture: Neuroscience, embodiment, and the future of design*. The MIT Press.

Shields, J., et al. (2016). Evaluating the veridicality of two-dimensional representations of three-dimensional architectural space through physiological response. In M. Couceiro da Costa et al. (Eds.), *Architectural research addressing societal challenges* (pp. 1001–1014). Taylor & Francis Group.

Taylor, J. L. (2009). Proprioception. In L. R. Squire (Ed.), *Encyclopedia of neuroscience* (pp. 1143–1149). Springer.

Torralba, A. (2005). Contextual influences on saliency. In L. Itti et al. (Eds.), *Neurobiology of attention* (pp. 586–592). Elsevier/Academic.

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), 766–786.

Varela, F. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies, 3*(4), 330–349.

Von Uexküll, J. (1992). A stroll through the worlds of animals and men: A picture book of invisible worlds. *Semiotica, 89*(4), 319–391. https://doi.org/10.1515/semi.1992.89.4.319

Weber, R. (1995). *On the aesthetics of architecture: A psychological approach to the structure and the order of perceived architectural space*. Avebury.

Weber, R., et al. (2002). The impact of formal properties on eye movement during the perception of architecture. *Journal of Architectural Planning and Research, 19*(1), 57–68.

Wollheim, R. (1973). *On art and the mind: Essays and lectures*. Allen Lane.

Woodruff Smith, D. (2013). *Husserl*. Taylor & Francis Group.

Yarbus, A., & Haigh, B. (1973). *Eye movements and vision*. Plenum Press.

Yoshimi, J. (2016). Prospects for a naturalized phenomenology. In D. O. Dahlstrom et al. (Eds.), *Philosophy of mind and phenomenology: Conceptual and empirical approaches* (pp. 299–321). Routledge.

Zumthor, P. (2010). *Thinking architecture* (3rd expanded ed.). Birkhäuser.