frontiers in Computational Neuroscience

REVIEW
published: 16 April 2020
doi: 10.3389/fncom.2020.00029

Attention in Psychology, Neuroscience, and Machine Learning

*Grace W. Lindsay**

Gatsby Computational Neuroscience Unit, Sainsbury Wellcome Centre, University College London, London, United Kingdom

Attention is the important ability to flexibly control limited computational resources. It has been studied in conjunction with many other topics in neuroscience and psychology including awareness, vigilance, saliency, executive control, and learning. It has also recently been applied in several domains in machine learning. The relationship between the study of biological attention and its use as a tool to enhance artificial neural networks is not always clear. This review starts by providing an overview of how attention is conceptualized in the neuroscience and psychology literature. It then covers several use cases of attention in machine learning, indicating their biological counterparts where they exist. Finally, the ways in which artificial attention can be further inspired by biology for the production of complex and integrative systems is explored.

Keywords: attention, artificial neural networks, machine learning, vision, memory, awareness

OPEN ACCESS

Edited by:
Adam Henry Marblestone,
Harvard University, United States

Reviewed by:
Mattia Rigotti,
IBM Research, United States
Mariya Toneva,
Carnegie Mellon University,
United States
H. Steven Scholte,
University of Amsterdam, Netherlands

***Correspondence:**
Grace W. Lindsay
gracewlindsay@gmail.com

Received: *02 December 2019*
Accepted: *23 April 2020*
Published: *16 April 2020*

Citation:
Lindsay GW (2020) Attention in Psychology, Neuroscience, and Machine Learning.
Front. Comput. Neurosci. 14:29.
doi: 10.3389/fncom.2020.00029

1. INTRODUCTION

Attention is a topic widely discussed publicly and widely studied scientifically. It has many definitions within and across multiple fields including psychology, neuroscience, and, most recently, machine learning (Chun et al., 2011; Cho et al., 2015). As William James wrote at the dawn of experimental psychology, "Everyone knows what attention is. It is the taking possession by the mind, in clear, and vivid form, of one out of what seems several simultaneously possible objects or trains of thought." Since James wrote this, many attempts have been made to more precisely define and quantify this process while also identifying the underlying mental and neural architectures that give rise to it. The glut of different experimental approaches and conceptualizations to study what is spoken of as a single concept, however, has led to something of a backlash amongst researchers. As was claimed in the title of a recent article arguing for a more evolution-informed approach to the concept, "No one knows what attention is" (Hommel et al., 2019).

Attention is certainly far from a clear or unified concept. Yet despite its many, vague, and sometimes conflicting definitions, there is a core quality of attention that is demonstrably of high importance to information processing in the brain and, increasingly, artificial systems. Attention is the flexible control of limited computational resources. Why those resources are limited and how they can best be controlled will vary across use cases, but the ability to dynamically alter and route the flow of information has clear benefits for the adaptiveness of any system.

The realization that attention plays many roles in the brain makes its addition to artificial neural networks unsurprising. Artificial neural networks are parallel processing systems comprised of individual units designed to mimic the basic input-output function of neurons. These models are currently dominating the machine learning and artificial intelligence (AI) literature. Initially constructed without attention, various mechanisms for dynamically re-configuring the representations or structures of these networks have now been added.

The following section, section 2, will cover broadly the different uses of the word attention in neuroscience and psychology, along with its connection to other common neuroscientific topics. Throughout, the conceptualization of attention as a way to control limited resources will be highlighted. Behavioral studies will be used to demonstrate the abilities and limits of attention while neural mechanisms point to the physical means through which these behavioral effects are manifested. In section 3, the state of attention research in machine learning will be summarized and relationships between artificial and biological attention will be indicated where they exist. And in section 4 additional ways in which findings from biological attention can influence its artificial counterpart will be presented.

The primary aim of this review is to give researchers in the field of AI or machine learning an understanding of how attention is conceptualized and studied in neuroscience and psychology in order to facilitate further inspiration where fruitful. A secondary aim is to inform those who study biological attention how these processes are being operationalized in artificial systems as it may influence thinking about the functional implications of biological findings.

2. ATTENTION IN NEUROSCIENCE AND PSYCHOLOGY

The scientific study of attention began in psychology, where careful behavioral experimentation can give rise to precise demonstrations of the tendencies and abilities of attention in different circumstances. Cognitive science and cognitive psychology aim to turn these observations into models of how mental processes could create such behavioral patterns. Many word models and computational models have been created that posit different underlying mechanisms (Driver, 2001; Borji and Itti, 2012).

The influence of single-cell neurophysiology in non-human primates along with non-invasive means of monitoring human brain activity such as EEG, fMRI, and MEG have made direct observation of the underlying neural processes possible. From this, computational models of neural circuits have been built that can replicate certain features of the neural responses that relate to attention (Shipp, 2004).

In the following sub-sections, the behavioral and neural findings of several different broad classes of attention will be discussed.

2.1. Attention as Arousal, Alertness, or Vigilance

In its most generic form, attention could be described as merely an overall level of alertness or ability to engage with surroundings. In this way it interacts with arousal and the sleep-wake spectrum. Vigilance in psychology refers to the ability to sustain attention and is therefore related as well. Note, while the use of these words clusters around the same meaning, they are sometimes used more specifically in different niche literature (Oken et al., 2006).

Studying subjects in different phases of the sleep-wake cycle, under sleep deprivation, or while on sedatives offers a view of how this form of attention can vary and what the behavioral consequences are. By giving subjects repetitive tasks that require a level of sustained attention—such as keeping a ball within a certain region on a screen—researchers have observed extended periods of poor performance in drowsy patients that correlate with changes in EEG signals (Makeig et al., 2000). Yet, there are ways in which tasks can be made more engaging that can lead to higher performance even in drowsy or sedated states. This includes increasing the promise of reward for performing the task, adding novelty or irregularity, or introducing stress (Oken et al., 2006). Therefore, general attention appears to have limited reserves that won't be deployed in the case of a mundane or insufficiently rewarding task but can be called upon for more promising or interesting work.

Interestingly, more arousal is not always beneficial. The Yerkes-Dodson curve (**Figure 1B**) is an inverted-U that represents performance as a function of alertness on sufficiently challenging tasks: at low levels of alertness performance is poor, at medium levels it is good, and at high levels it becomes poor again. The original study used electric shocks in mice to vary the level of alertness, but the finding has been repeated with other measures (Diamond, 2005). It may explain why psychostimulants such as Adderall or caffeine can work to increase focus in some people at some doses but become detrimental for others (Wood et al., 2014).

The neural circuits underlying the sleep-wake cycle are primarily in the brain stem (Coenen, 1998). These circuits control the flow of information into the thalamus and then onto cortex. Additionally, neuromodulatory systems play a large role in the control of generalized attention. Norepinephrine, acetylcholine, and dopamine are believed to influence alertnesss, orienting to important information, and executive control of attention, respectively (Posner, 2008). The anatomy of neuromodulators matches their function as well. Neurons that release norepinephrine, for example, have their cell bodies in the brain stem but project very broadly across the brain, allowing them to control information processing broadly (**Figure 1A**).

2.2. Sensory Attention

In addition to overall levels of arousal and alertness, attention can also be selectively deployed by an awake subject to specific sensory inputs. Studying attention within the context of a specific sensory system allows for tight control over both stimuli and the locus of attention. Generally, to look for this type of attention the task used needs to be quite challenging. For example, in a change detection task, the to-be-detected difference between two stimuli may be very slight. More generally, task difficulty can be achieved by presenting the stimulus for only a very short period of time or only very weakly.

A large portion of the study of attention in systems neuroscience and psychology centers on visual attention in particular (Kanwisher and Wojciulik, 2000). This may reflect the general trend in these fields to emphasis the study of visual processing over other sensory systems (Hutmacher, 2019), along with the dominant role vision plays in the primate brain.

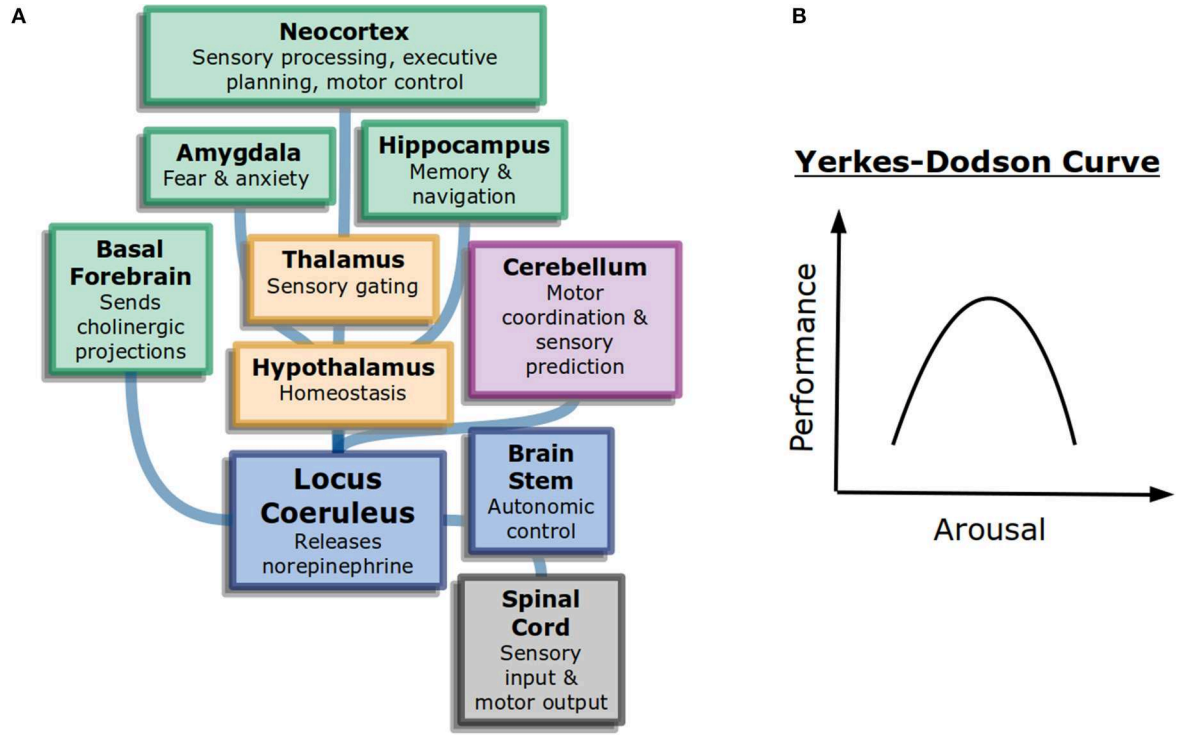

FIGURE 1 | General attention and alertness **(A)** Cells in the locus coeruleus release norepinephrine (also known as noradrenaline) onto many parts of the brain with different functions, including onto other neuromodulatory systems. This contributes to overall arousal (Samuels and Szabadi, 2008). Colors here represent different divisions of the brain: forebrain (green), diencephalon (yellow), and brainstem (blue). **(B)** The Yerkes-Dodson curve describes the nonlinear relationship between arousal and performance on challenging tasks.

Furthermore, visual stimuli are frequently used in studies meant to address more general, cognitive aspects of attention as well.

Visual attention can be broken down broadly into spatial and feature-based attention.

2.2.1. Visual Spatial Attention

Saccades are small and rapid eye movements made several times each second. As the fovea offers the highest visual resolution on the retina, choosing where to place it is essentially a choice about where to deploy limited computational resources. In this way, eye movements indicate the locus of attention. As this shift of attention is outwardly visible it is known as overt visual attention.

By tracking eye movements as subjects are presented with different images, researchers have identified image patterns that automatically attract attention. Such patterns are defined by oriented edges, spatial frequency, color contrast, intensity, or motion (Itti and Koch, 2001). Image regions that attract attention are considered "salient" and are computed in a "bottom-up" fashion. That is, they don't require conscious or effortful processing to identify and are likely the result of built-in feature detectors in the visual system. As such, saliency can be computed very quickly. Furthermore, different subjects tend to agree on which regions are salient, especially those identified in the first few saccades (Tatler et al., 2005).

Salient regions can be studied in "free-viewing" situations, that is, when the subject is not given any specific instructions about how to view the image. When a particular task is assigned, the interplay between bottom-up and "top-down" attention becomes clear. For example, when instructed to saccade to a specific visual target out of an array, subjects may incorrectly saccade to a particularly salient distractor instead (van Zoest and Donk, 2005). More generally, task instructions can have a significant effect on the pattern of saccades generated when subjects are viewing a complex natural image and given high-level tasks (e.g., asked to assess the age of a person or guess their socio-economic status). Furthermore, the natural pattern of eye movements when subjects perform real world tasks, like sandwich making, can provide insights to underlying cognitive processes (Hayhoe and Ballard, 2005).

When subjects need to make multiple saccades in a row they tend not to return to locations they have recently attended and may be slow to respond if something relevant occurs there. This phenomenon is known as inhibition of return (Itti and Koch, 2001). Such behavior pushes the visual system to not just exploit image regions originally deemed most salient but to explore other areas as well. It also means the saccade generating system needs to have a form of memory; this is believed to be implemented by short-term inhibition of the representation of recently-attended locations.

While eye movements are an effective means of controlling visual attention, they are not the only option. "Covert" spatial attention is a way of emphasizing processing of different spatial

locations without an overt shift in fovea location. Generally, in the study of covert spatial attention, subjects must fixate on a central point throughout the task. They are cued to covertly attend to a location in their peripheral vision where stimuli relevant for their visual task will likely appear. For example, in an orientation discrimination task, after the spatial cue is provided an oriented grating will flash in the cued location and the subject will need to indicate its orientation. On invalidly-cued trials (when the stimulus appears in an uncued location), subjects perform worse than on validly-cued (or uncued) trials (Anton-Erxleben and Carrasco, 2013). This indicates that covert spatial attention is a limited resource that can be flexibly deployed and aids in the processing of visual information.

Covert spatial attention is selective in the sense that certain regions are selected for further processing at the expense of others. This has been referred to as the "spotlight" of attention. Importantly, for covert—as opposed to overt—attention the input to the visual system can be identical while the processing of that input is flexibly selective.

Covert spatial attention can be impacted by bottom-up saliency as well. If an irrelevant but salient object is flashed at a location that then goes on to have a task relevant stimulus, the exogenous spatial attention drawn by the irrelevant stimulus can get applied to the task relevant stimulus, possibly providing a performance benefit. If it is flashed at an irrelevant location, however, it will not help, and can harm performance (Berger et al., 2005). Bottom-up/exogenous attention has a quick time course, impacting covert attention for 80–130 ms after the distractor appears (Anton-Erxleben and Carrasco, 2013).

In some theories of attention, covert spatial attention exists to help guide overt attention. Particularly, the pre-motor theory of attention posits that the same neural circuits plan saccades and control covert spatial attention (Rizzolatti et al., 1987). The frontal eye field (FEF) is known to be involved in the control of eye movements. Stimulating the neurons in FEF at levels too low to evoke eye movements has been shown to create effects similar to covert attention (Moore et al., 2003). In this way, covert attention may be a means of deciding where to overtly look. The ability to covertly attend may additionally be helpful in social species, as eye movements convey information about knowledge and intent that may best be kept secret (Klein et al., 2009).

To study the neural correlates of covert spatial attention, researchers identify which aspects of neural activity differ based only on differences in the attentional cue (and not on differences in bottom-up features of the stimuli). On trials where attention is cued toward the receptive field of a recorded neuron, many changes in the neural activity have been observed (Noudoost et al., 2010; Maunsell, 2015). A commonly reported finding is an increase in firing rates, typically of 20–30% (Mitchell et al., 2007). However, the exact magnitude of the change depends on the cortical area studied, with later areas showing stronger changes (Luck et al., 1997; Noudoost et al., 2010). Attention is also known to impact the variability of neural firing. In particular, it decreases trial-to-trial variability as measured via the Fano Factor and decreases noise correlations between pairs of neurons. Attention has even been found to impact the electrophysiological properties of neurons in a way that reduces their likelihood of firing in bursts and also decreases the height of individual action potentials (Anderson et al., 2013).

In general, the changes associated with attention are believed to increase the signal-to-noise ratio of the neurons that represent the attended stimulus, however they can also impact communication between brain areas. To this end, attention's effect on neural synchrony is important. Within a visual area, attention has been shown to increase spiking coherence in the gamma band—that is at frequencies between 30 and 70 Hz (Fries et al., 2008). When a group of neurons fires synchronously, their ability to influence shared downstream areas is enhanced. Furthermore, attention may also be working to directly coordinate communication across areas. Synchronous activity between two visual areas can be a sign of increased communication and attention has been shown to increase synchrony between the neurons that represent the attended stimulus in areas V1 and V4, for example (Bosman et al., 2012). Control of this cross-area synchronization appears to be carried out by the pulvinar (Saalmann et al., 2012).

In addition to investigating how attention impacts neurons in the visual pathways, studies have also searched for the source of top-down attention (Noudoost et al., 2010; Miller and Buschman, 2014). The processing of bottom-up attention appears to culminate with a saliency map produced in the lateral intraparietal area (LIP). The cells here respond when salient stimuli are in their receptive field, including task-irrelevant but salient distractors. Prefrontal areas such as FEF, on the other hand, appear to house the signals needed for top-down control of spatial attention and are less responsive to distractors.

While much of the work on the neural correlates of sensory attention focuses on the cortex, subcortical areas appear to play a strong role in the control and performance benefits of attention as well. In particular, the superior colliculus assists in both covert and overt spatial attention and inactivation of this region can impair attention (Krauzlis et al., 2013). And, as mentioned above, the pulvinar plays a role in attention, particularly with respect to gating effects on cortex (Zhou et al., 2016).

2.2.2. Visual Feature Attention

Feature attention is another form of covert selective attention. In the study of feature attention, instead of being cued to attend to a particular location, subjects are cued on each trial to attend to a particular visual feature such as a specific color, a particular shape, or a certain orientation. The goal of the task may be to detect if the cued feature is present on the screen or readout another one of its qualities (e.g., to answer "what color is the square?" should result in attention first deployed to squares). Valid cueing about the attended feature enhances performance. For example, when attention was directed toward a particular orientation, subjects were better able to detect faint gratings of that orientation than of any other orientation (Rossi and Paradiso, 1995). While the overall task (e.g., detection of an oriented grating) remains the same, the specific instructions (detection of 90° grating vs. 60° vs. 30°) will be cued on each individual trial, or possibly blockwise. Successful trial-wise cueing indicates that this form of attention can be flexibly deployed on fast timescales.

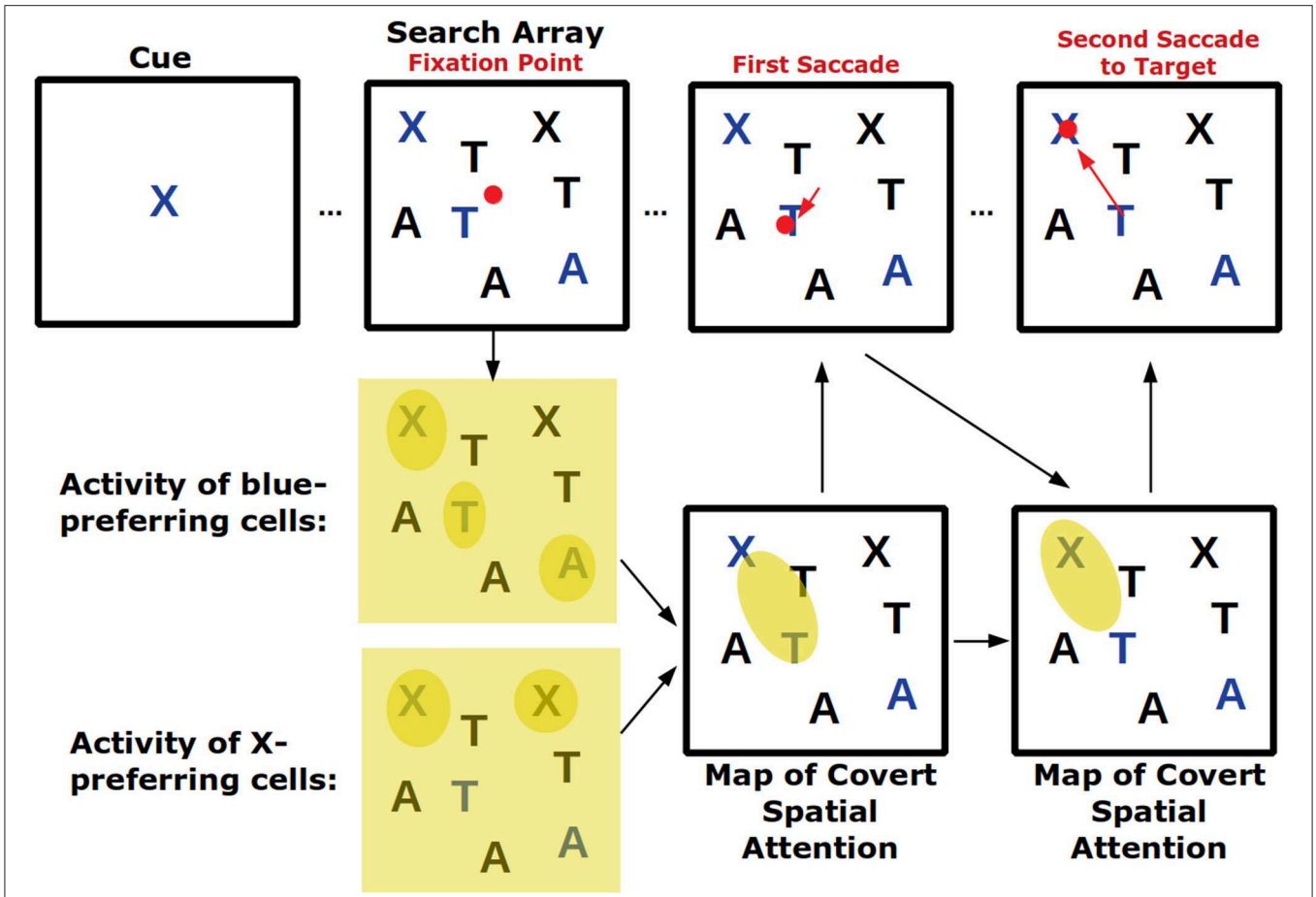

FIGURE 2 | Visual search tasks engage many forms of visual attention. Across the top row the progression of a visual search task is shown. First, a cue indicates the target of the visual search, in this case a blue X. Then a search array appears with many non-targets. Top-down feature attention to cells that represent the color blue and the shape X will increase their firing throughout the visual field but firing will be strongest where blue or Xs actually occur. These neural response will play a role in generating a map of covert spatial attention which can be used to explore visual space before saccading. After the shift in overt attention with the first saccade, the covert attention map is remade. Finally, the target is located and successfully saccaded to. If the visual array contained a pop-out stimulus (for example a green O) it may have captured covert spatial attention in a bottom-up way and led to an additional incorrect saccade.

Visual search tasks are also believed to activate feature-based attention (**Figure 2**). In these tasks, an array of stimuli appears on a screen and subjects need to indicate—frequently with an eye movement—the location of the cued stimulus. As subjects are usually allowed to make saccades throughout the task as they search for the cued stimulus, this task combines covert feature-based attention with overt attention. In fact, signals of top-down feature-based attention have been found in FEF, the area involved in saccade choice (Zhou and Desimone, 2011). Because certain features can create a pop-out effect—for example, a single red shape amongst several black ones will immediately draw attention—visual search tasks also engage bottom-up attention which, depending on the task, may need to be suppressed (Wolfe and Horowitz, 2004).

Neural effects of feature-based attention in the visual system are generally similar to those of spatial attention. Neurons that represent the attended feature, for example, have increased firing rates, and those that represent very different features have suppressed rates (Treue and Trujillo, 1999). As opposed to spatial attention, however, feature-based attention is spatially-global. This means that when deploying attention to a particular feature the activity of the neurons that represent that feature anywhere in visual space are modulated (Saenz et al., 2002). Another difference between spatial and feature attention is the question of how sources of top-down attention target the correct neurons in the visual system. The retinotopic map, wherein nearby cells represent nearby spatial locations, makes spatial targeting straightforward, but cells are not as neatly organized according to preferred visual features.

The effects of spatial and feature attention appear to be additive (Hayden and Gallant, 2009). Furthermore, both feature and spatial attention are believed to create their effects by acting on the local neural circuits that implement divisive normalization in visual cortex (Reynolds and Heeger, 2009). Modeling work has shown that many of the neural effects of selective attention can be captured by assuming that top-down

connections provide targeted synaptic inputs to cells in these circuits (Lindsay et al., 2019). However, models that rely on effects of the neuromodulator acetylcholine can also replicate neural correlates of attention (Sajedin et al., 2019).

Potential sources of top-down feature-based attention have been found in prefrontal cortex where sustained activity encodes the attended feature (Bichot et al., 2015; Paneri and Gregoriou, 2017). Inactivating the ventral prearcuate area impairs performance on search tasks. From prefrontal areas, attention signals are believed to travel in a reverse hierarchical way wherein higher visual areas send inputs to those below them (Ahissar and Hochstein, 2000).

A closely related topic to feature attention is object attention. Here, attention is not deployed to an abstract feature in advance of a visual stimulus, but rather it is applied to a particular object in the visual scene (Chen, 2012). The initial feedforward pass of activity through the visual hierarchy is able to pre-attentively segregate objects from their backgrounds in parallel across the visual field, provided these objects have stark and salient differences from the background. In more crowded or complex visual scenes, recurrent and serial processing is needed in order to identify different objects (Lamme and Roelfsema, 2000). Serial processing involves moving limited attentional resources from one location in the image to another; it can take the form of shifts in either covert or overt spatial attention (Buschman and Miller, 2009). Recurrent connections in the visual system—that is, both horizontal connections from nearby neurons in the same visual area and feedback connections from those in higher visual areas—aid in figure-ground segregation and object identification. The question of how the brain performs perceptual grouping of low-level features into a coherent object identity has been studied for nearly a century. It is believed that attention may be required for grouping, particularly for novel or complex objects (Roelfsema and Houtkamp, 2011). This may be especially important in visual search tasks that require locating an object that is defined by a conjunction of several features.

Neurally, the effects of object-based attention can spread slowly through space as parts of an object are mentally traced (Roelfsema et al., 1998). Switching attention to a location outside an object appears to incur a greater cost than switching to the same distance away but within the object (Brown and Denney, 2007). In addition, once attention is applied to a visual object, it is believed to activate feature-based attention for the different features of that object across the visual field (O'Craven et al., 1999).

Another form of attention sometimes referred to as feature attention involves attending to an entire feature dimension. An example of this is the Stroop test, wherein the names of colors are written in different colored ink and subjects either need to read the word itself or say the color of the ink. Here attention cannot be deployed to a specific feature in advance, only to the dimensions word or color. Neurally, the switch between dimensions appears to impact sensory coding in the visual stream and is controlled by frontal areas (Liu et al., 2003).

2.2.3. Computational Models of Visual Attention

Visual attention, being one of the most heavily-studied topics in the neuroscience of attention, has inspired many computational models of how attention works. In general, these models synthesize various neurophysiological findings in order to help explain how the behavioral impacts of attention arise (Heinke and Humphreys, 2005).

Several computational models meant to calculate saliency have been devised (Itti and Koch, 2001). These models use low-level visual feature detectors—usually designed to match those in the visual system—to create an image-specific saliency map that can predict the saccade patterns of humans in response to the same image. Another approach to calculating saliency based on information theoretic first principles has also been explored and was able to account for certain visual search behaviors (Bruce and Tsotsos, 2009).

Some of the behavioral and neural correlates of attention are similar whether the attention is bottom-up or top-down. In the Biased Competition Model of attention, stimuli compete against each other to dominate the neural response (Desimone, 1998). Attention (bottom-up or top-down) can thus work by biasing this competition toward the stimulus that is the target of attention. While the Biased Competition Model is sometimes used simply as a "word model" to guide intuition, explicit computational instantiations of it have also been built. A hierarchical model of the visual pathway that included top-down biasing as well as local competition mediated through horizontal connections was able to replicate multiple neural effects of attention (Deco and Rolls, 2004). A model embodying similar principles but using spiking neurons was also implemented (Deco and Rolls, 2005).

Similar models have been constructed explicitly to deal with attribute naming tasks such as the Stroop test described above. The Selective Attention Model (SLAM), for example, has local competition in both the sensory encoding and motor output modules and can mimic known properties of response times in easier and more challenging Stroop-like tests (Phaf et al., 1990).

Visual perception has been framed and modeled as a problem of Bayesian inference (Lee and Mumford, 2003). Within this context, attention can help resolve uncertainty under settings where inference is more challenging, typically by modulating priors (Rao, 2005). For example, in Chikkerur et al. (2010) spatial attention functions to reduce uncertainty about object identity and feature attention reduces spatial uncertainty. These principles can capture both behavioral and neural features of attention and can be implemented in a biologically-inspired neural model.

The feature similarity gain model of attention (FSGM) is a description of the neural effects of top-down attention that can be applied in both the feature and spatial domain (Treue and Trujillo, 1999). It says that the way in which a neuron's response is modulated by attention depends on that neuron's tuning. Tuning is a description of how a neuron responds to different stimuli, so according to the FSGM a neuron that prefers (that is, responds strongly to), e.g., the color blue, will have its activity enhanced by top-down attention to blue. The FSGM also says attention to non-preferred stimuli will

cause a decrease in firing and that, whether increased or decreased, activity is scaled multiplicatively by attention. Though not initially defined as a computational model, this form of neural modulation has since been shown through modeling to be effective at enhancing performance on challenging visual tasks (Lindsay and Miller, 2018).

Other models conceptualize attention as a dynamic routing of information through a network. An implementation of this form of attention can be found in the Selective Attention for Identification Model (SAIM) (Heinke and Humphreys, 2003). Here, attention routes information from the retina to a representation deemed the "focus of attention"; depending on the current task, different parts of the retinal representation will be mapped to the focus of attention.

2.2.4. Attention in Other Sensory Modalities

A famous example of the need for selective attention in audition is the "cocktail party problem": the difficulty of focusing on the speech from one speaker in a crowded room of multiple speakers and other noises (Bronkhorst, 2015). Solving the problem is believed to involve "early" selection wherein low level features of a voice such as pitch are used to determine which auditory information is passed on for further linguistic processing. Interestingly, selective auditory attention has the ability to control neural activity at even the earliest level of auditory processing, the cochlea (Fritz et al., 2007).

Spatial and feature attention have also been explored in the somatosensory system. Subjects cued to expect a tap at different parts on their body are better able to detect the sensation when that cue is valid. However, these effects seem weaker than they are in the visual system (Johansen-Berg and Lloyd, 2000). Reaction times are faster in a detection task when subjects are cued about the orientation of a stimulus on their finger (Schweisfurth et al., 2014).

In a study that tested subjects' ability to detect a taste they had been cued for it was shown that validly-cued tastes can be detected at lower concentrations than invalidly-cued ones (Marks and Wheeler, 1998). This mimics the behavioral effects found with feature-based visual attention. Attention to olfactory features has not been thoroughly explored, though visually-induced expectations about a scent can aid its detection (Gottfried and Dolan, 2003; Keller, 2011).

Attention can also be spread across modalities to perform tasks that require integration of multiple sensory signals. In general, the use of multiple congruent sensory signals aids detection of objects when compared to relying only on a single modality. Interestingly, some studies suggest that humans may have a bias for the visual domain, even when the signal from another domain is equally valid (Spence, 2009). Specifically, the visual domain appears to dominate most in tasks that require identifying the spatial location of a cue (Bertelson and Aschersleben, 1998). This can be seen most readily in ventriloquism, where the visual cue of the dummy's mouth moving overrides auditory evidence about the true location of the vocal source. Visual evidence can also override tactile evidence, for example, in the context of the rubber arm illusion (Botvinick and Cohen, 1998).

Another effect of the cross-modal nature of sensory processing is that an attentional cue in one modality can cause an orienting of attention in another modality (Spence and Driver, 2004). Generally, the attention effects in the non-cued modality are weaker. This cross-modal interaction can occur in the context of both endogenous ("top-down") and exogenous ("bottom-up") attention.

2.3. Attention and Executive Control

With multiple simultaneous competing tasks, a central controller is needed to decide which to engage in and when. What's more, how to best execute tasks can depend on history and context. Combining sensory inputs with past knowledge in order to coordinate multiple systems for the job of efficient task selection and execution is the role of executive control, and this control is usually associated with the prefrontal cortex (Miller and Buschman, 2014). As mentioned above, sources of top-down visual attention have also been located in prefrontal regions. Attention can reasonably be thought of as the output of executive control. The executive control system must thus select the targets of attention and communicate that to the systems responsible for implementing it. According to the reverse hierarchy theory described above, higher areas signal to those from which they get input which send the signal on to those below them and so on (Ahissar and Hochstein, 2000). This means that, at each point, the instructions for attention must be transformed into a representation that makes sense for the targeted region. Through this process, the high level goals of the executive control region can lead to very specific changes, for example, in early sensory processing.

Executive control and working memory are also intertwined, as the ability to make use of past information as well as to keep a current goal in mind requires working memory. Furthermore, working memory is frequently identified as sustained activity in prefrontal areas. A consequence of the three-way relationship between executive control, working memory, and attention is that the contents of working memory can impact attention, even when not desirable for the task (Soto et al., 2008). For example, if a subject has to keep an object in working memory while simultaneously performing a visual search for a separate object, the presence of the stored object in the search array can negatively interfere with the search (Soto et al., 2005). This suggests that working memory can interfere with the executive control of attention. However, there still appears to be additional elements of that control that working memory alone does not disrupt. This can be seen in studies wherein visual search performance is even worse when subjects believe they will need to report the memorized item but are shown a search array for the attended item instead (Olivers and Eimer, 2011). This suggests that, while all objects in working memory may have some influence over attention, the executive controller can choose which will have the most.

Beyond the flexible control of attention within a sensory modality, attention can also be shifted between modalities. Behavioral experiments indicate that switching attention either between two different tasks within a sensory modality (for example, going from locating a visual object to identifying it) or

between sensory modalities (switching from an auditory task to a visual one) incurs a computational cost (Pashler, 2000). This cost is usually measured as the extent to which performance is worse on trials just after the task has been switched vs. those where the same task is being repeated. Interestingly, task switching within a modality seems to incur a larger cost than switching between modalities (Murray et al., 2009). A similar result is found when switching between or across modes of response (for example, pressing a bottom vs. verbal report), suggesting this is not specific to sensory processing (Arrington et al., 2003). Such findings are believed to stem from the fact that switching within a modality requires a reconfiguration of the same neural circuits, which is more difficult than merely engaging the circuitry of a different sensory system. An efficient executive controller would need to be aware of these costs when deciding to shift attention and ideally try to minimize them; it has been shown that switch costs can be reduced with training (Gopher, 1996).

The final question regarding the executive control of attention is how it evolves with learning. Eye movement studies indicate that searched-for items can be detected more rapidly in familiar settings rather than novel ones, suggesting that previously-learned associations guide overt attention (Chun and Jiang, 1998). Such benefits are believed to rely on the hippocampus (Aly and Turk-Browne, 2017). In general, however, learning how to direct attention is not as studied as other aspects of the attention process. Some studies have shown that subjects can enhance their ability to suppress irrelevant task information, and the generality of that suppression depends on the training procedure (Kelley and Yantis, 2009). Looking at the neural correlates of attention learning, imaging results suggest that the neural changes associated with learning do not occur in the sensory pathways themselves but rather in areas more associated with attentional control (Kelley and Yantis, 2010). Though not always easy to study, the development of attentional systems in infancy and childhood may provide further clues as to how attention can be learned (Reynolds and Romano, 2016).

2.4. Attention and Memory

Attention and memory have many possible forms of interaction. If memory has a limited capacity, for example, it makes sense for the brain to be selective about what is allowed to enter it. In this way, the ability of attention to dynamically select a subset of total information is well-matched to the needs of the memory system. In the other direction, deciding to recall a specific memory is a choice about how to deploy limited resources. Therefore, both memory encoding and retrieval can rely on attention.

The role of attention in memory encoding appears quite strong (Aly and Turk-Browne, 2017). For information to be properly encoded into memory, it is best for it be the target of attention. When subjects are asked to memorize a list of words while simultaneously engaging in a secondary task that divides their attention, their ability to consciously recall those words later is impaired (though their ability to recognize the words as familiar is not so affected) (Gardiner and Parkin, 1990). Imaging studies have shown that increasing the difficulty of the secondary task weakens the pattern of activity related to memory encoding in the left ventral inferior frontal gyrus and anterior hippocampus and increases the representation of secondary task information in dorsolateral prefrontal and superior parietal regions (Uncapher and Rugg, 2005). Therefore, without the limited neural processing power placed on the task of encoding, memory suffers. Attention has also been implicated in the encoding of spatially-defined memories and appears to stabilize the representations of place cells (Muzzio et al., 2009).

Implicit statistical learning can also be biased by attention. For example, in Turk-Browne et al. (2005) subjects watched a stream of stimuli comprised of red and green shapes. The task was to detect when a shape of the attended color appeared twice in a row. Unbeknownst to the subjects, certain statistical regularities existed in the stream such that there were triplets of shapes likely to occur close together. When shown two sets of three shapes—one an actual co-occurring triplet and another a random selection of shapes of the same color—subjects recognized the real triplet as more familiar, but only if the triplets were from the attended color. The statistical regularities of the unattended shapes were not learned.

Yet some learning can occur even without conscious attention. For example, in Watanabe (2003) patients engaged in a letter detection task located centrally in their visual field while random dot motion was shown in the background at sub-threshold contrast. The motion had 10% coherence in a direction that was correlated with the currently-presented letter. Before and after learning this task, subjects performed an above-threshold direction classification task. After learning the task, direction classification improved only for the direction associated with the targeted letters. This suggests a reward-related signal activated by the target led to learning about a non-attended component of the stimulus.

Many behavioral studies have explored the extent to which attention is needed for memory retrieval. For example, by asking subjects to simultaneously recall a list of previously-memorized words and engage in a secondary task like card sorting, researchers can determine if memory retrieval pulls from the same limited pool of attentional resources as the task. Some such studies have found that retrieval is impaired by the co-occurrence of an attention-demanding task, suggesting it is an attention-dependent process. The exact findings, however, depend on the details of the memory and non-memory tasks used (Lozito and Mulligan, 2006).

Even if memory retrieval does not pull from shared attentional resources, it is still clear that some memories are selected for more vivid retrieval at any given moment than others. Therefore, a selection process must occur. An examination of neuroimaging results suggests that the same parietal brain regions responsible for the top-down allocation and bottom-up capture of attention may play analogous roles during memory retrieval (Wagner et al., 2005; Ciaramelli et al., 2008).

Studies of memory retrieval usually look at medium to long-term memory but a mechanism for attention to items in working memory has also been proposed (Manohar et al., 2019). It relies on two different mechanisms of working memory: synaptic traces for non-attended items and sustained activity for the attended one.

Some forms of memory occur automatically and within the sensory processing stream itself. Priming is a well-known phenomenon in psychology wherein the presence of a stimulus at one point in time impacts how later stimuli are processed or interpreted. For example, the word "doctor" may be recognized more quickly following the word "hospital" than the word "school." In this way, priming requires a form of implicit memory to allow previous stimuli to impact current ones. Several studies on conceptual or semantic priming indicate that attention to the first stimulus is required for priming effects to occur (Ballesteros and Mayas, 2015); this mirrors findings that attention is required for memory encoding more generally.

Most priming is positive, meaning that the presence of a stimulus at one time makes the detection and processing of it or a related stimulus more likely at a later time. In this way, priming can be thought of as biasing bottom-up attention. However, top-down attention can also create negative priming. In negative priming, when stimuli that functioned as a distractor on the previous trial serve as the target of attention on the current trial, performance suffers (Frings et al., 2015). This may stem from a holdover effect wherein the mechanisms of distractor suppression are still activated for the now-target stimulus.

Adaptation can also be considered a form of implicit memory. Here, neural responses decrease after repeated exposure to the same stimulus. By reducing the response to repetition, changes in the stimulus become more salient. Attention—by increasing the neural response to attended stimuli—counters the effects of adaptation (Pestilli et al., 2007; Anton-Erxleben et al., 2013). Thus, both with priming and adaptation, top-down attention can overcome automatic processes that occur at lower levels which may be guiding bottom-up attention.

3. ATTENTION IN MACHINE LEARNING

While the concept of artificial attention has come up prior to the current resurgence of artificial neural networks, many of its popular uses today center on ANNs (Mancas et al., 2016). The use of attention mechanisms in artificial neural networks came about—much like the apparent need for attention in the brain—as a means of making neural systems more flexible. Attention mechanisms in machine learning allow a single trained artificial neural network to perform well on multiple tasks or tasks with inputs of variable length, size, or structure. While the spirit of attention in machine learning is certainly inspired by psychology, its implementations do not always track with what is known about biological attention, as will be noted below.

In the form of attention originally developed for ANNs, attention mechanisms worked within an encoder-decoder framework and in the context of sequence models (Cho et al., 2015; Chaudhari et al., 2019). Specifically, an input sequence will be passed through an encoder (likely a recurrent neural network) and the job of the decoder (also likely a recurrent neural network) will be to output another sequence. Connecting the encoder and decoder is an attention mechanism.

Commonly, the output of the encoder is a set of a vectors, one for each element in the input sequence. Attention helps determine which of these vectors should be used to generate the output. Because the output sequence is dynamically generated one element at a time, attention can dynamically highlight different encoded vectors at each time point. This allows the decoder to flexibly utilize the most relevant parts of the input sequence.

The specific job of the attention mechanism is to produce a set of scalar weightings, α_t^i, one for each of the encoded vectors (v^i). At each step t, the attention mechanism (ϕ) will take in information about the decoder's previous hidden state (h_{t-1}) and the encoded vectors to produce unnormalized weightings:

$$\tilde{\boldsymbol{\alpha}}_t = \boldsymbol{\phi}(h_{t-1}, v) \tag{1}$$

Because attention is a limited resource, these weightings need to represent relative importance. To ensure that the α values sum to one, the unnormalized weightings are passed through a softmax:

$$\alpha_t^i = \frac{\exp(\tilde{\alpha}_t^i)}{\sum_j \exp(\tilde{\alpha}_t^j)} \tag{2}$$

These attention values scale the encoded vectors to create a single context vector on which the decoder can be conditioned:

$$c_t = \sum_j \alpha_t^j v^j \tag{3}$$

This form of attention can be made entirely differentiable and so the whole network can be trained end-to-end with simple gradient descent.

This type of artificial attention is thus a form of iterative re-weighting. Specifically, it dynamically highlights different components of a pre-processed input as they are needed for output generation. This makes it flexible and context dependent, like biological attention. As such it is also inherently dynamic. While sequence modeling already has an implied temporal component, this form of attention can also be applied to static inputs and outputs (as will be discussed below in the context of image processing) and will thus introduce dynamics into the model.

In the traditional encoder-decoder framework without attention, the encoder produced a fixed-length vector that was independent of the length or features of the input and static during the course of decoding. This forced long sequences or sequences with complex structure to be represented with the same dimensionality as shorter or simpler ones and didn't allow the decoder to interrogate different parts of the input during the decoding process. But encoding the input as a set of vectors equal in length to the input sequence makes it possible for the decoder to selectively attend to the portion of the input sequence relevant at each time point of the decoding. Again, as in interpretations of attention in the brain, attention in artificial systems is helpful as a way to flexibly wield limited resources. The decoder can't reasonably be conditioned on the entirety of the input so at some point a bottleneck must be introduced. In the system without attention, the fixed-length encoding vector was a bottleneck. When an attention mechanism is added, the encoding can be

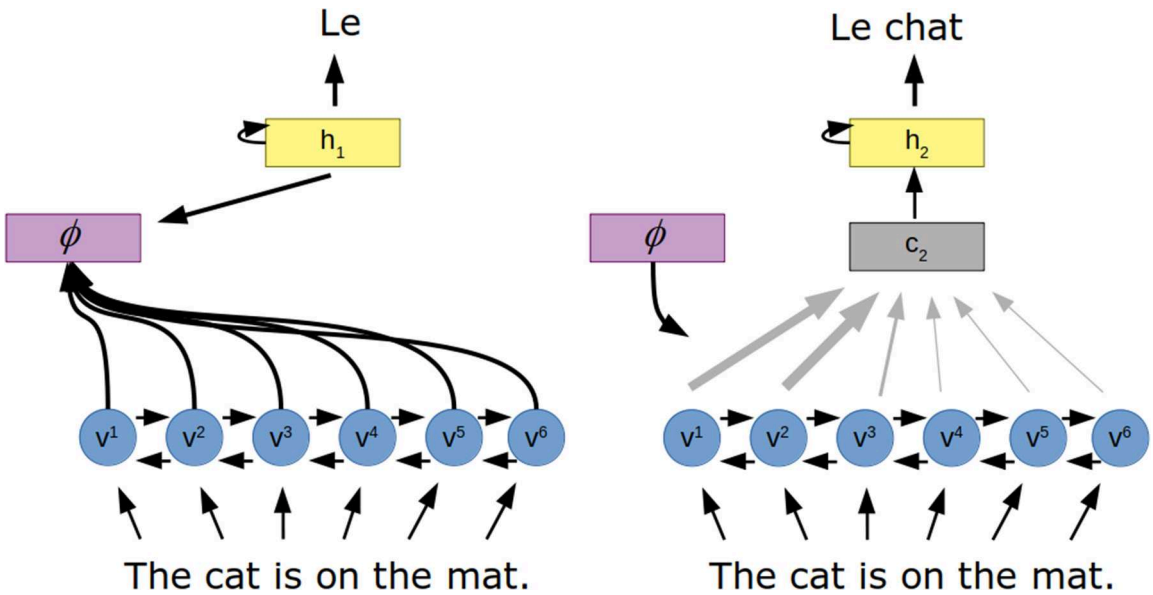

FIGURE 3 | Attention for neural machine translation. The to-be-translated sentence is encoded to a series of vectors (v) via a recurrent neural network. The attention mechanism (ϕ) uses the hidden state of the decoder (h) and these vectors to determine how the encoded vectors should be combined to produce a context vector (c), which influences the next hidden state of the decoder and thus the next word in the translated sentence.

larger because the bottleneck (in the form of the context vector) will be produced dynamically as the decoder determines which part of the input to attend to.

The motivation for adding such attention mechanisms to artificial systems is of course to improve their performance. But another claimed benefit of attention is interpretability. By identifying on which portions of the input attention is placed (that is, which α^i values are high) during the decoding process, it may be possible to gain an understanding of why the decoder produced the output that it did. However, caution should be applied when interpreting the outputs of attention as they may not always explain the behavior of the model as expected (Jain and Wallace, 2019; Wiegreffe and Pinter, 2019).

In the following subsections, specific applications of this general attention concept will be discussed, along with some that don't fit neatly into this framework. Further analogies to the biology will also be highlighted.

3.1. Attention for Natural Language Processing

As described above, attention mechanisms have frequently been added to models charged with processing sequences. Natural language processing (NLP) is one of the most common areas of application for sequence modeling. And, though it was not the original domain of attention in machine learning—nor does it have the most in common with biology—NLP is also one of the most common areas of application for attention (Galassi et al., 2019).

An early application of the this form of attention in artificial neural networks was to the task of translation (Bahdanau et al., 2014) (**Figure 3**). In this work, a recurrent neural network encodes the input sentence as a set of "annotation" vectors, one for each word in the sentence. The output, a sentence in the target language, is generated one word at a time by a recurrent neural network. The probability of each generated word is a function of the previously generated word, the hidden state of the recurrent neural network and a context vector generated by the attention mechanism. Here, the attention mechanism is a small feedforward neural network that takes in the hidden state of the output network as well as the current annotation vector to create the weighting over all annotation vectors.

Blending information from all the words in the sentence this way allows the network to pull from earlier or later parts when generating an output word. This can be especially useful for translating between languages with different standard word orders. By visualizing the locations in the input sentence to which attention was applied the authors observed attention helping with this problem.

Since this initial application, many variants of attention networks for language translation have been developed. In Firat et al. (2016), the attention mechanism was adapted so it could be used to translate between multiple pairs of languages rather than just one. In Luong et al. (2015), the authors explore different structures of attention to determine if the ability to access all input words at once is necessary. And in Cheng et al. (2016), attention mechanisms were added to the recurrent neural networks that perform the sentence encoding and decoding in order to more flexibly create sentence representations.

In 2017, the influential "Attention is All You Need" paper utilized a very different style of architecture for machine translation (Vaswani et al., 2017). This model doesn't have any recurrence, making it simpler to train. Instead, words in the

sentence are encoded in parallel and these encodings generate key and query representations that are combined to create attention weightings. These weightings scale the word encodings themselves to create the next layer in the model, a process known as "self-attention." This process repeats, and eventually interacts with the autoregressive decoder which also has attention mechanisms that allow it to flexibly focus on the encoded input (as in the standard form of attention) and on the previously generated output. The Transformer—the name given to this new attention architecture—outperformed many previous models and quickly became the standard for machine translation as well as other tasks (Devlin et al., 2018).

Interestingly, self-attention has less in common with biological attention than the recurrent attention models originally used for machine translation. First, it reduces the role of recurrence and dynamics, whereas the brain necessarily relies on recurrence in sequential processing tasks, including language processing and attentional selection. Second, self-attention provides a form of horizontal interaction between words—which allows for words in the encoded sentence to be processed in the context of those around them—but this mechanism does not include an obvious top-down component driven by the needs of the decoder. In fact, self-attention has been shown under certain circumstances to simply implement a convolution, a standard feedforward computation frequently used in image processing (Andreoli, 2019; Cordonnier et al., 2019). In this way, self-attention is more about creating a good encoding than performing a task-specific attention-like selection based on limited resources. In the context of a temporal task, its closest analogue in psychology may be priming because priming alters the encoding of subsequent stimuli based on those that came before. It is of course not the direct goal of machine learning engineers to replicate the brain, but rather to create networks that can be easily trained to perform well on tasks. These different constraints mean that even large advances in machine learning do not necessarily create more brain-like models.

While the study of attention in human language processing is not as large as other areas of neuroscience research, some work has been done to track eye movements while reading (Myachykov and Posner, 2005). They find that people will look back at previous sections of text in order to clarify what they are currently reading, particularly in the context of finding the antecedent of a pronoun. Such shifts in overt attention indicate what previous information is most relevant for the current processing demands.

3.2. Attention for Visual Tasks

As in neuroscience and psychology, a large portion of studies in machine learning are done on visual tasks. One of the original attention-inspired tools of computer vision is the saliency map, which identifies which regions in an image are most salient based on a set of low-level visual features such as edges, color, or depth and how they differ from their surround (Itti and Koch, 2001). In this way, saliency maps indicate which regions would be captured by "bottom-up" attention in humans and animals. Computer scientists have used saliency maps as part of their image processing pipeline to identify regions for further processing.

In more recent years, computer vision models have been dominated by deep learning. And since their success in the 2012 ImageNet Challenge (Russakovsky et al., 2015), convolutional neural networks have become the default architecture for visual tasks in machine learning.

The architecture of convolutional neural networks is loosely based on the mammalian visual system (Lindsay, 2020). At each layer, a bank of filters is applied to the activity of the layer below (in the first layer this is the image). This creates a $H \times W \times C$ tensor of neural activity with the number of channels, C equal to the number of filters applied and H and W representing the height and width of the 2-D feature maps that result from the application of a filter.

Attention in convolutional neural networks has been used to enhance performance on a variety of tasks including classification, segmentation, and image-inspired natural language processing. Also, as in the neuroscience literature, these attentional processes can be divided into spatial and feature-based attention.

3.2.1. Spatial Attention

Building off of the structures used for attention in NLP tasks, visual attention has been applied to image captioning. In Xu et al. (2015), the encoding model is a convolutional neural network. The attention mechanism works over the activity at the fourth convolutional layer. As each word of the caption is generated, a different pattern of weighting across spatial locations of the image representation is created. In this way, attention for caption generation replaces the set of encoded word vectors in a translation task with a set of encoded image locations. Visualizing the locations with high weights, the model appears to attend to the object most relevant to the current word being generated for the caption.

This style of attention is referred to as "soft" because it produces a weighted combination of the visual features over spatial locations (**Figure 4B**). "Hard" attention is an alternative form that chooses a single spatial location to be passed into the decoder at the expense of all others (**Figure 4A**). In Xu et al. (2015), to decide which location should receive this hard attention, the attention weights generated for each spatial location were treated as probabilities. One location is chosen according to these probabilities. Adding this stochastic element to the network makes training more difficult, yet it was found to perform somewhat better than soft attention.

A 2014 study used reinforcement learning to train a hard attention network to perform object recognition in challenging conditions (Mnih et al., 2014). The core of this model is a recurrent neural network that both keeps track of information taken in over multiple "glimpses" made by the network and outputs the location of the next glimpse. For each glimpse, the network receives a fovea-like input (central areas are represented with high resolution and peripheral with lower) from a small patch of the image. The network has to integrate the information gained from these glimpses to find and classify the object in the image. This is similar to the hard attention described above, except the selection of a location here determines which part of the image is sampled next (whereas in the case above

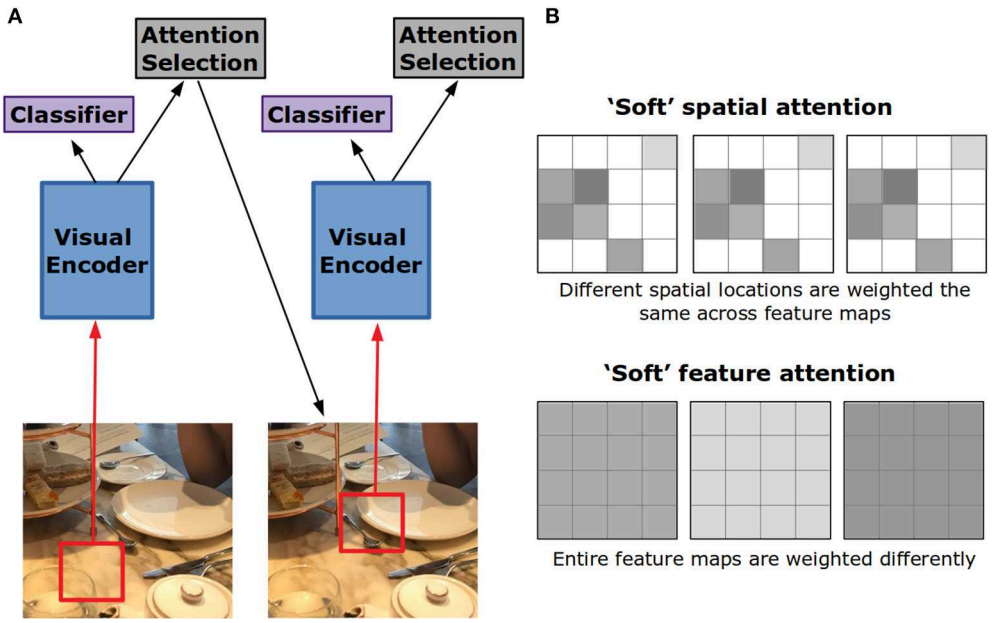

FIGURE 4 | Hard vs. soft visual attention in artificial neural networks. **(A)** In hard attention, the network only gets input from a small portion of the whole image. This portion is iteratively chosen by the network through an attention selection mechanism. If the input is foveated, the network can use the lower resolution periphery to guide this selection. **(B)** Feature maps in convolutional neural networks are 2-D grids of activation created by the application of a filter to the layer below. In soft spatial attention, different locations on these grids are weighted differently. In soft feature attention, different feature maps are weighted differently.

it determined which of the already-processed image locations would be passed to the decoder). With the use of these glimpses, the network is not required to process all of the image, saving computational resources. It can also help when multiple objects are present in the image and the network must classify each (Ba et al., 2014). Recent work has shown that adding a pre-training step enhances the performance of hard attention applied to complex images (Elsayed et al., 2019).

In many ways, the correspondence between biological and artificial attention is strongest when it comes to visual spatial attention. For example, this form of hard attention—where different locations of the image are sequentially-sampled for further processing—replicates the process of saccading and is therefore akin to overt visual attention in the neuroscience and psychology literature. Insofar as soft attention dynamically re-weights different regions of the network's representation of the image without any change in the input to the network, it is akin to covert spatial attention. Also, as the mode of application for soft attention involves multiplicative scaling of the activity of all units at a specific location, it replicates neural findings about covert spatial attention.

Soft spatial attention has been used for other tasks, including visual question and answering (Chen et al., 2015; Xu and Saenko, 2016; Yang et al., 2016) and action recognition in videos (Sharma et al., 2015). Hard attention has also been used for instance segmentation (Ren and Zemel, 2017) and for fine-grained classification when applied using different levels of image resolution (Fu et al., 2017).

3.2.2. Feature Attention

In the case of soft spatial attention, weights are different in different spatial locations of the image representation yet they are the same across all feature channels at that location. That is, the activity of units in the network representing different visual features will all be modified the same way if they represent the same location in image space. Feature attention makes it possible to dynamically re-weight individual feature maps, creating a spatially global change in feature processing.

In Stollenga et al. (2014), a convolutional neural network is equipped with a feature-based attention mechanism. After an image is passed through the standard feedforward architecture, the activity of the network is passed into a policy that determines how the different feature maps at different layers should be weighted. This re-weighting leads to different network activity which leads to different re-weightings. After the network has run for several timesteps the activity at the final layer is used to classify the object in the image. The policy that determines the weighting values is learned through reinforcement learning, and can be added to any pre-trained convolutional neural network.

The model in Chen et al. (2017) combines feature and spatial attention to aid in image captioning. The activity of the feedforward pass of the convolutional network is passed into the attention mechanism along with the previously generated word to create attention weightings for different channels at each layer in the CNN. These weights are used to scale activity and then a separate attention mechanism does the same procedure for generating spatial weightings. Both spatial and feature attention

weights are generated and applied to the network at each time point.

In the model in De Vries et al. (2017), the content of a question is used to control how a CNN processes an image for the task of visual question and answering. Specifically, the activity of a language embedding network is passed through a multi-layer perceptron to produce the additive and multiplicative parameters for batch normalization of each channel in the CNN. This procedure, termed conditional batch normalization, functions as a form of question-dependent feature attention.

A different form of dynamic feature re-weighting appears in "squeeze-and-excitation" networks (Hu et al., 2018). In this architecture, the weightings applied to different channels are a nonlinear function of the activity of the other channels at the same layer. As with "self-attention" described above, this differs in spirit from more "top-down" approaches where weightings are a function of activity later in the network and/or biased by the needs of the output generator. Biologically speaking, this form of interaction is most similar to horizontal connections within a visual area, which are known to carry out computations such as divisive normalization (Carandini and Heeger, 2012).

In the study of the biology of feature-based attention, subjects are usually cued to attend to or search for specific visual features. In this way, the to-be-attended features are known in advance and relate to the specific sub-task at hand (e.g., detection of a specific shape on a given trial of a general shape detection task). This differs from the above instances of artificial feature attention, wherein no external cue biases the network processing before knowledge about the specific image is available. Rather, the feature re-weighting is a function of the image itself and meant to enhance the performance of the network on a constant task (note this was also the case for the forms of artificial spatial attention described).

The reason for using a cueing paradigm in studies of biological attention is that it allows the experimenter to control (and thus know) where attention is placed. Yet, it is clear that even without explicit cueing, our brains make decisions about where to place attention constantly; these are likely mediated by local and long-range feedback connections to the visual system (Wyatte et al., 2014). Therefore, while the task structure differs between the study of biological feature attention and its use in artificial systems, this difference may only be superficial. Essentially, the artificial systems are using feedforward image information to internally generate top-down attentional signals rather than being given the top-down information in the form of a cue.

That being said, some artificial systems do allow for externally-cued feature attention. For example setting a prior over categories in the network in Cao et al. (2015) makes it better at localizing the specific category. The network in Wang et al. (2014), though not convolutional, has a means of biasing the detection of specific object categories as well. And in Lindsay and Miller (2018), several performance and neural aspects of biological feature attention during a cued object detection task were replicated using a CNN. In Luo et al. (2020), the costs and benefits of using a form of cued attention in CNNs were explored.

As mentioned above, the use of multiplicative scaling of activity is in line with certain findings from biological visual attention. Furthermore, modulating entire feature maps by the same scalar value is aligned with the finding mentioned above that feature attention acts in a spatially global way in the visual system.

3.3. Multi-Task Attention

Multi-task learning is a challenging topic in machine learning. When one network is asked to perform several different tasks—for example, a CNN that must classify objects, detect edges, and identify salient regions—training can be difficult as the weights needed to do each individual task may contradict each other. One option is have a set of task-specific parameters that modulate the activity of the shared network differently for each task. While not always called it, this can reasonably be considered a form of attention, as it flexibly alters the functioning of the network.

In Maninis et al. (2019), a shared feedforward network is trained on all of multiple tasks, while task specific skip connections and squeeze-and-excitation blocks are trained to modulate this activity only on their specific task. This lets the network benefit from sharing processing that is common to all tasks while still specializing somewhat to each.

A similar procedure was used in Rebuffi et al. (2017) to create a network that performs classification on multiple different image domains. There, the domain could be identified from the input image making it possible to select the set of task-specific parameters automatically at run-time.

In Zhao et al. (2018), the same image can be passed into the network and be classified along different dimensions (e.g. whether the person in the picture is smiling or not, young or old). Task-specific re-weighting of feature channels is used to execute these different classifications.

The model in Strezoski et al. (2019) uses what could be interpreted as a form of hard feature attention to route information differently in different tasks. Binary masks over feature channels are chosen randomly for each task. These masks are applied in a task-specific way during training on all tasks and at run-time. Note that in this network no task-specific attentional parameters are learned, as these masks are pre-determined and fixed during training. Instead, the network learns to use the different resulting information pathways to perform different tasks.

In a recent work, the notion of task-specific parameters was done away with entirely (Levi and Ullman, 2020). Instead, the activations of a feedforward CNN are combined with a task input and passed through a second CNN to generate a full set of modulatory weights. These weights then scale the activity of the original network in a unit-specific way (thus implementing both spatial and feature attention). The result is a single set of feedforward weights capable of flexibly engaging in multiple visual tasks.

When the same input is processed differently according to many different tasks, these networks are essentially implementing a form of within-modality task switching that relies on feature attention. In this way, it is perhaps most similar to the Stroop test described previously.

3.4. Attention to Memory

Deep neural networks tend not to have explicit memory, and therefore attention to memory is not studied. Neural Turing Machines, however, are a hybrid neural architecture that includes external memory stores (Graves et al., 2014). The network, through training, learns how to effectively interact with these stores to perform tasks such as sorting and repetition of stored sequences. Facilitating this interaction is a form of attention. Memories are stored as a set of vectors. To retrieve information from this store, the network generates a weight for each vector and calculates a weighted sum of the memories. To determine these weights, a recurrent neural network (which receives external and task-relevant input) outputs a vector and memories are weighted in accordance to their similarity to this vector. Thus, at each point in time, the network is able to access context-relevant memories.

As described previously, how the brain chooses what memories to attend to and then attends to them is not entirely clear. The use of a similarity metric in this model means that memories are retrieved based on their overlap with a produced activity vector, similar to associative memory models in the neuroscience literature. This offers a mechanism for the latter question—that is, how attention to memory could be implemented in the brain. The activity vector that the model produces controls what memories get attended and the relationship with biology is less clear here.

4. IDEAS FOR FUTURE INTERACTION BETWEEN ARTIFICIAL AND BIOLOGICAL ATTENTION

As has been shown, some amount of inspiration from biology has already led to several instances of attention in artificial neural networks (summarized in **Figure 5**). While the addition of such attention mechanisms has led to appreciable increases in performance in these systems, there are clearly still many ways in which they fall short and additional opportunities for further inspiration exist. In the near term, this inspiration will likely be in the form of incremental improvements to specialized artificial systems as exist now. However, the true promise of brain-inspired AI should deliver a more integrated, multiple-purpose agent that can engage flexibly in many tasks.

4.1. How to Enhance Performance

There are two components to the study of how attention works in the brain that can be considered flip sides of the same coin. The first is the question of how attention enhances performance in the way that it does—that is, how do the neural changes associated with attention make the brain better at performing tasks. The second is how and why attention is deployed in the way that it is—what factors lead to the selection of certain items or tasks for attention and not others.

Neuroscientists have spent a lot of time investigating the former question. In large part, the applicability of these findings to artificial neural systems, however, may not be straightforward. Multiplicative scaling of activity appears in both biological and artificial systems and is an effective means of implementing attention. However, many of the observed effects of attention in the brain make sense mainly as a means of increasing the signal carried by noisy, spiking neurons. This includes increased synchronization across neurons and decreased firing variability. Without analogs for these changes in deep neural networks, it is hard to take inspiration from them. What's more, the training procedures for neural networks can automatically determine the changes in activity needed to enhance performance on a well-defined task and so lessons from biological changes may not be as relevant.

On the other hand, the observation that attention can impact spiking-specific features such as action potential height, burstiness, and precise spike times may indicate the usefulness of spiking networks. Specifically, spiking models offer more degrees of freedom for attention to control and thus allow attention to possibly have larger and/or more nuanced impacts.

Looking at the anatomy of attention may provide usable insights to people designing architectures for artificial systems. For example, visual attention appears to modulate activity more strongly in later visual areas like V4 (Noudoost et al., 2010), whereas auditory attention can modulate activity much earlier in the processing stream. The level at which attention should act could thus be a relevant architectural variable. In this vein, recent work has shown that removing self-attention from the early layers of a Transformer model enhances its performance on certain natural language processing tasks and also makes the model a better predictor of human fMRI signals during language processing (Toneva and Wehbe, 2019).

The existence of cross-modal cueing—wherein attention cued in one sensory modality can cause attention to be deployed to the same object or location in another modality—indicates some amount of direct interaction between different sensory systems. Whereas many multi-modal models in machine learning use entirely separate processing streams that are only combined at the end, allowing some horizontal connections between different input streams may help coordinate their processing.

Attention also interacts with the kind of adaptation that normally occurs in sensory processing. Generally, neural network models do not have mechanisms for adaptation—that is, neurons have no means of reducing their activity if given the same input for multiple time steps. Given that adaptation helps make changes and anomalies stand out, it may be useful to include. In a model with adaption, attention mechanisms should work to reactivate adapted neurons if the repeated stimulus is deemed important.

Finally, some forms of attention appear to act in multiple ways on the same system. For example, visual attention is believed to both: (1) enhance the sensitivity of visual neurons in the cortex by modulating their activity and (2) change subcortical activity such that sensory information is readout differently (Birman and Gardner, 2019; Sreenivasan and Sridharan, 2019). In this way, attention uses two different mechanisms, in different parts of the brain, to create its effect. Allowing attention to modulate multiple components of a model architecture in complementary ways may allow it to have more robust and effective impacts.

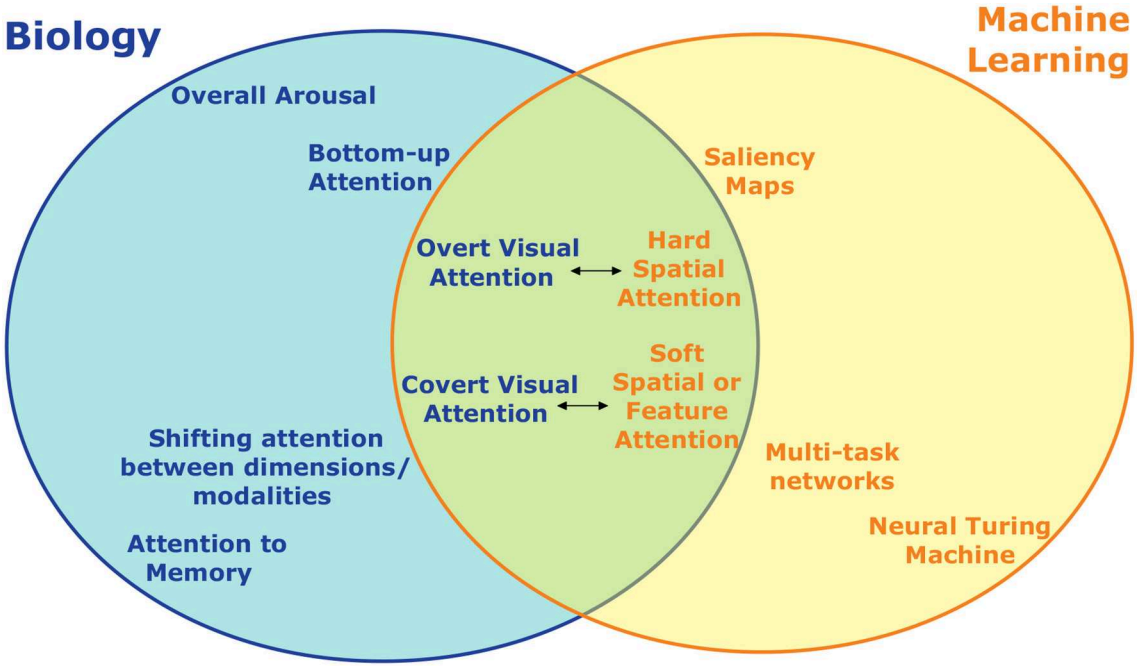

FIGURE 5 | An incomplete summary of the different types of attention studied in neuroscience/psychology and machine learning and how they relate. On the left are divisions of attention studied biologically, on the right are those developed for artificial intelligence and machine learning. Topics at the same horizontal location are to some extent analogous, with the distance between them indicating how close the analogy is. Forms of visual attention, for example, have the most overlap and are the most directly comparable across biology and machine learning. Some forms of attention, such as overall arousal, don't have an obvious artificial analogue.

4.2. How to Deploy Attention

The question of how to deploy attention is likely the more relevant challenge for producing complex and integrated artificial intelligence. Choosing the relevant information in a stream of incoming stimuli, picking the best task to engage in, or deciding whether to engage in anything at all requires that an agent have an integrative understanding of its state, environment, and needs.

The most direct way to take influence from biological attention is to mimic it directly. Scanpath models, for example, have existed in the study of saliency for many years. They attempt to predict the series of fixations that humans make while viewing images (Borji and Itti, 2019). A more direct approach to training attention was used in Linsley et al. (2018). Here, a large dataset of human top-down attention was collected by having subjects label the regions of images most relevant for object classification. The task-specific saliency maps created through this method were used to train attention in a deep convolutional neural network whose main task was object recognition. They found that influencing the activity of intermediate layers with this method could increase performance. Another way of learning a teacher's saliency map was given in Zagoruyko and Komodakis (2016).

Combined training on tasks and neural data collected from human visual areas has also helped the performance of CNNs (Fong et al., 2018). Using neural data collected during attention tasks in particular could help train attention models. Such transfer could also be done for other tasks. For example, tracking eye movements during reading could inform NLP models; thus far, eye movements have been used to help train a part-of-speech tagging model (Barrett et al., 2016). Interestingly, infants may learn from attending to what adults around them attend to and the coordination of attention more broadly across agents may be very helpful in a social species. Therefore, the attention of others should influence how attention is guided. Attempts to coordinate joint attention will need to be integrated into attention systems (Kaplan and Hafner, 2006; Klein et al., 2009).

Interestingly, infants may learn from attending to what adults around them attend to and the coordination of attention more broadly across agents may be very helpful in a social species. Therefore, the attention of others should influence how attention is guided. Attempts to coordinate joint attention will need to be integrated into attention systems (Kaplan and Hafner, 2006; Klein et al., 2009). Activities would likely need to flexibly decide which of several possible goals should be achieved at any time and therefore where attention should be placed. This problem clearly interacts closely with issues around reinforcement learning—particularly hierarchical reinforcement learning which involves the choosing of subtasks—as such decisions must be based on expected positive or negative outcomes. Indeed, there is a close relationship between attention and reward as previously-rewarded stimuli attract attention even in contexts where they no longer provide reward (Camara et al., 2013). A better understanding of how humans choose which tasks to engage in and when should allow human behavior to inform the design of a multi-task AI.

To this end, the theory put forth in Shenhav et al. (2013), which says that allocation of the brain's limited ability to control

different processes is based on the expected value of that control, may be of use. In this framework, the dorsal anterior cingulate cortex is responsible for integrating diverse information—including the cognitive costs of control—in order to calculate the expected value of control and thus direct processes like attention. Another approach for understanding human executive control in complex tasks is inverse reinforcement learning. This method was recently applied to a dataset of eye movements during visual search in order to determine the reward functions and policies used by humans (Zelinsky et al., 2020).

An additional factor that drives biological attention but is perhaps underrepresented in artificial attention systems is curiosity (Gottlieb et al., 2013). In biology, novel, confusing, and surprising stimuli can grab attention, and inferotemporal and perirhinal cortex are believed to signal novel visual situations via an adaptation mechanism that reduces responses to familiar inputs. Reinforcement learning algorithms that include novelty as part of the estimate of the value of a state can encourage this kind of exploration (Jaegle et al., 2019). How exactly to calculate surprise or novelty in different circumstances is not always clear, however. Previous work on biological attention has understood attention selection in Bayesian terms of surprise or information gathering and these framings may be useful for artificial systems (Itti and Baldi, 2006 ; Mirza et al., 2019).

A final issue in the selection of attention is how conflicts are resolved. Given the brain's multiple forms of attention—arousal, bottom-up, top-down, etc.—how do conflicts regarding the appropriate locus of attention get settled? Looking at the visual system, it seems that the local circuits that these multiple systems target are burdened with this task. These circuits receive neuromodulatory input along with top-down signals which they must integrate with the bottom-up input driving their activity. Horizontal connections mediate this competition, potentially using winner-take-all mechanisms. This can be mimicked in the architecture of artificial systems.

4.3. Attention and Learning

Attention, through its role in determining what enters memory, guides learning. Most artificial systems with attention include the attention mechanism throughout training. In this way, the attention mechanism is trained along with the base architecture; however, with the exception of the Neural Turing Machine, the model does not continue learning once the functioning attention system is in place. Therefore, the ability of attention to control learning and memory is still not explicitly considered in these systems.

Attention could help make efficient use of data by directing learning to the relevant components and relationships in the input. For example, saliency maps have been used as part of the pre-processing for various computer vision tasks (Lee et al., 2004; Wolf et al., 2007; Bai and Wang, 2014). Focusing subsequent processing only on regions that are intrinsically salient can prevent wasteful processing on irrelevant regions and, in the context of network training, could also prevent overfitting to these regions. Using saliency maps in this way, however, requires a definition of saliency that works for the problem at hand. Using the features of images that capture bottom-up attention in humans has worked for some computer vision problems; looking at human data in other modalities may be useful as well.

In a related vein, studies on infants suggest that they have priors that guide their attention to relevant stimuli such as faces. Using such priors could bootstrap learning both of how to process important stimuli and how to better attend to their relevant features (Johnson, 2001).

In addition to deciding which portions of the data to process, top-down attention can also be thought of as selecting which elements of the network should be most engaged during processing. Insofar as learning will occur most strongly in the parts of the network that are most engaged, this is another means by which attention guides learning. Constraining the number of parameters that will be updated in response to any given input is an effective form of regularization, as can be seen in the use of dropout and batch normalization. Attention—rather than randomly choosing which units to engage and disengage—is constrained to choose units that will also help performance on this task. It is therefore a more task-specific form of regularization.

In this way, attention may be particularly helpful for continual learning where the aim is to update a network to perform better on a specific task while not disrupting performance on the other tasks the network has already learned to do. A related concept, conditional computation, has recently been applied to the problem of continual learning (Lin et al., 2019). In conditional computation, the parameters of a network are a function of the current input (it can thus be thought of as an extreme form of the type of modulation done by attention); optimizing the network for efficient continual learning involves controlling the amount of interference between different inputs. More generically, it may be helpful to think of attention, in part, as a means of guarding against undesirable synaptic changes.

Attention and learning also work in a loop. Specifically, attention guides what is learned about the world and internal world models are used to guide attention. This inter-dependency has recently been formalized in terms of a reinforcement learning framework that also incorporates cognitive Bayesian inference models that have succeeded in explaining human learning and decision making (Radulescu et al., 2019). Interconnections between basal ganglia and prefrontal cortex are believed to support the interplay between reinforcement learning and attention selection.

At a more abstract level, the mere presence of attention in the brain's architecture can influence representation learning. The global workspace theory of consciousness says that at any moment a limited amount of information selected from the brain's activity can enter working memory and be available for further joint processing (Baars, 2005). Inspired by this, the 'consciousness prior' in machine learning emphasizes a neural network architecture with a low-dimensional representation that arises from attention applied to an underlying high-dimensional state representation (Bengio, 2017). This low-D representation should efficiently represent the world at an abstract level such that it can be used to summarize and make predictions about future states. The presence of this attention-mediated bottleneck has a trickle-down effect that encourages disentangled representations

at all levels such that they can be flexibly combined to guide actions and make predictions.

Conscious attention is required for the learning of many complex skills such as playing a musical instrument. However once fully learned, these processes can become automatic, possibly freeing attention up to focus on other things (Treisman et al., 1992). The mechanisms of this transformation are not entirely clear but insofar as they seem to rely on moving the burden of the task to different, possibly lower/more reflexive brain areas, it may benefit artificial systems to have multiple redundant pathways that can be engaged differently by attention (Poldrack et al., 2005).

4.4. Limitations of Attention: Bugs or Features?

Biological attention does not work perfectly. As mentioned above, performance can suffer when switching between different kinds of attention, arousal levels need be just right in order to reach peak performance, and top-down attention can be interrupted by irrelevant but salient stimuli. A question when transferring attention to artificial systems is are these limitations bugs to be avoided or features to be incorporated?

Distractability, in general, seems like a feature of attention rather than a bug. Even when attempting to focus on a task it is beneficial to still be aware of—and distractable by—potentially life-threatening changes in the environment. The problem comes only when an agent is overly distractable to inputs that do not pose a threat or provide relevant information. Thus, artificial systems should balance the strength of top down attention such that it still allows for the processing of unexpected but informative stimuli. For example, attentional blink refers to the phenomenon wherein a subject misses a second target in a stream of targets and distractors if it occurs quickly after a first target (Shapiro et al., 1997). While this makes performance worse, it may be necessary to give the brain time to process and act on the first target. In this way, it prevents distractability to ensure follow through.

Any agent, artificial or biological, will have some limitations on its energy resources. Therefore, prudent decisions about when to engage in the world versus enter an energy-saving state such as sleep will always be of relevance. For many animals sleep occurs according to a schedule but, as was discussed, it can also be delayed or interrupted by attention-demanding situations. The decision about when to enter a sleep state must thus be made based on a cost-benefit analysis of what can be gained by staying awake. Because sleep is also known to consolidate memories and perform other vital tasks beyond just energy conservation, this decision may be a complex one. Artificial systems will need to have an integrative understanding of their current state and future demands to make this decision.

5. CONCLUSIONS

Attention is a large and complex topic that sprawls across psychology, neuroscience, and artificial intelligence. While many of the topics studied under this name are non-overlapping in their mechanisms, they do share a core theme of the flexible control of limited resources. General findings about flexibility and wise uses of resources can help guide the development of AI, as can specific findings about the best means of deploying attention to specific sensory modalities or tasks.

AUTHOR CONTRIBUTIONS

GL conceived and wrote the article and generated the figures.

FUNDING

This work was supported by a Marie Skłodowska-Curie Individual Fellowship (No. 844003) and a Sainsbury Wellcome Centre/Gatsby Computational Unit Fellowship.

ACKNOWLEDGMENTS

The author would like to thank Jacqueline Gottlieb and the three reviewers for their insights and pointers to references.

REFERENCES

Ahissar, M., and Hochstein, S. (2000). The spread of attention and learning in feature search: effects of target distribution and task difficulty. *Vis. Res.* 40, 1349–1364. doi: 10.1016/S0042-6989(00)00002-X

Aly, M., and Turk-Browne, N. B. (2017). "How hippocampal memory shapes, and is shaped by, attention," in *The Hippocampus From Cells to Systems*, eds D. E. Hannula and M. C. Duff (Cham: Springer), 369–403.

Anderson, E. B., Mitchell, J. F., and Reynolds, J. H. (2013). Attention-dependent reductions in burstiness and action-potential height in macaque area V4. *Nat. Neurosci.* 16, 1125–1131. doi: 10.1038/nn.3463

Andreoli, J.-M. (2019). Convolution, attention and structure embedding. *arXiv [preprint]*. arXiv: 1905.01289.

Anton-Erxleben, K., and Carrasco, M. (2013). Attentional enhancement of spatial resolution: linking behavioural and neurophysiological evidence. *Nat. Rev. Neurosci.* 14, 188–200. doi: 10.1038/nrn3443

Anton-Erxleben, K., Herrmann, K., and Carrasco, M. (2013). Independent effects of adaptation and attention on perceived speed. *Psychol. Sci.* 24, 150–159. doi: 10.1177/0956797612449178

Arrington, C. M., Altmann, E. M., and Carr, T. H. (2003). Tasks of a feather flock together: Similarity effects in task switching. *Mem. Cogn.* 31, 781–789. doi: 10.3758/BF03196116

Ba, J., Mnih, V., and Kavukcuoglu, K. (2014). Multiple object recognition with visual attention. *arXiv [preprint]*. arXiv:1412.7755.

Baars, B. J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Prog. Brain Res.* 150, 45–53. doi: 10.1016/S0079-6123(05)50004-9

Bahdanau, D., Cho, K., and Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv [preprint]*. arXiv:1409.0473.

Bai, X., and Wang, W. (2014). Saliency-SVM: an automatic approach for image segmentation. *Neurocomputing* 136, 243–255. doi: 10.1016/j.neucom.2014.01.008

Ballesteros, S., and Mayas, J. (2015). Selective attention affects conceptual object priming and recognition: a study with young and older adults. *Front. Psychol.* 5:1567. doi: 10.3389/fpsyg.2014.01567

Barrett, M., Bingel, J., Keller, F., and Søgaard, A. (2016). "Weakly supervised part-of-speech tagging using eye-tracking data," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (Berlin), 579–584.

Bengio, Y. (2017). The consciousness prior. *arXiv [preprint]*. arXiv:1709.08568.

Berger, A., Henik, A., and Rafal, R. (2005). Competition between endogenous and exogenous orienting of visual attention. *J. Exp. Psychol.* 134, 207–221. doi: 10.1037/0096-3445.134.2.207

Bertelson, P., and Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychon. Bull. Rev.* 5, 482–489. doi: 10.3758/BF03208826

Bichot, N. P., Heard, M. T., DeGennaro, E. M., and Desimone, R. (2015). A source for feature-based attention in the prefrontal cortex. *Neuron* 88, 832–844. doi: 10.1016/j.neuron.2015.10.001

Birman, D., and Gardner, J. L. (2019). A flexible readout mechanism of human sensory representations. *Nat. Commun.* 10, 1–13. doi: 10.1038/s41467-019-11448-7

Borji, A., and Itti, L. (2012). State-of-the-art in visual attention modeling. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 185–207. doi: 10.1109/TPAMI.2012.89

Borji, A., and Itti, L. (2019). Cat2000: a large scale fixation dataset for boosting saliency research. *arXiv [preprint]*. arXiv:1505.03581.

Bosman, C. A., Schoffelen, J.-M., Brunet, N., Oostenveld, R., Bastos, A. M., Womelsdorf, T., et al. (2012). Attentional stimulus selection through selective synchronization between monkey visual areas. *Neuron* 75, 875–888. doi: 10.1016/j.neuron.2012.06.037

Botvinick, M., and Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature* 391, 756–756. doi: 10.1038/35784

Bronkhorst, A. W. (2015). The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Attent. Percept. Psychophys.* 77, 1465–1487. doi: 10.3758/s13414-015-0882-9

Brown, J. M., and Denney, H. I. (2007). Shifting attention into and out of objects: evaluating the processes underlying the object advantage. *Percept. Psychophys.* 69, 606–618. doi: 10.3758/BF03193918

Bruce, N. D., and Tsotsos, J. K. (2009). Saliency, attention, and visual search: an information theoretic approach. *J. Vis.* 9:5. doi: 10.1167/9.3.5

Buschman, T. J., and Miller, E. K. (2009). Serial, covert shifts of attention during visual search are reflected by the frontal eye fields and correlated with population oscillations. *Neuron* 63, 386–396. doi: 10.1016/j.neuron.2009.06.020

Camara, E., Manohar, S., and Husain, M. (2013). Past rewards capture spatial attention and action choices. *Exp. Brain Res.* 230, 291–300. doi: 10.1007/s00221-013-3654-6

Cao, C., Liu, X., Yang, Y., Yu, Y., Wang, J., Wang, Z., et al. (2015). "Look and think twice: Capturing top-down visual attention with feedback convolutional neural networks," in *Proceedings of the IEEE International Conference on Computer Vision* (Santiago, CA), 2956–2964.

Carandini, M., and Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13, 51–62. doi: 10.1038/nrn3136

Chaudhari, S., Polatkan, G., Ramanath, R., and Mithal, V. (2019). An attentive survey of attention models. *arXiv [preprint]*. arXiv:1904.02874.

Chen, K., Wang, J., Chen, L.-C., Gao, H., Xu, W., and Nevatia, R. (2015). ABC-CNN: an attention based convolutional neural network for visual question answering. *arXiv [preprint]*. arXiv:1511.05960.

Chen, L., Zhang, H., Xiao, J., Nie, L., Shao, J., Liu, W., et al. (2017). "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI), 5659–5667. doi: 10.1109/CVPR.2017.667

Chen, Z. (2012). Object-based attention: a tutorial review. *Attent. Percept. Psychophys.* 74, 784–802. doi: 10.3758/s13414-012-0322-z

Cheng, J., Dong, L., and Lapata, M. (2016). Long short-term memory-networks for machine reading. *arXiv preprint arXiv:1601.06733*. doi: 10.18653/v1/D16-1053

Chikkerur, S., Serre, T., Tan, C., and Poggio, T. (2010). What and where: a bayesian inference theory of attention. *Vis. Res.* 50, 2233–2247. doi: 10.1016/j.visres.2010.05.013

Cho, K., Courville, A., and Bengio, Y. (2015). Describing multimedia content using attention-based encoder-decoder networks. *IEEE Trans. Multimed.* 17, 1875–1886. doi: 10.1109/TMM.2015.2477044

Chun, M. M., Golomb, J. D., and Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annu. Rev. Psychol.* 62, 73–101. doi: 10.1146/annurev.psych.093008.100427

Chun, M. M., and Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cogn. Psychol.* 36, 28–71. doi: 10.1006/cogp.1998.0681

Ciaramelli, E., Grady, C. L., and Moscovitch, M. (2008). Top-down and bottom-up attention to memory: a hypothesis (atom) on the role of the posterior parietal cortex in memory retrieval. *Neuropsychologia* 46, 1828–1851. doi: 10.1016/j.neuropsychologia.2008.03.022

Coenen, A. M. (1998). Neuronal phenomena associated with vigilance and consciousness: from cellular mechanisms to electroencephalographic patterns. *Conscious. Cogn.* 7, 42–53. doi: 10.1006/ccog.1997.0324

Cordonnier, J.-B., Loukas, A., and Jaggi, M. (2019). On the relationship between self-attention and convolutional layers. *arXiv [preprint]*. arXiv:1911.03584.

De Vries, H., Strub, F., Mary, J., Larochelle, H., Pietquin, O., and Courville, A. C. (2017). "Modulating early visual processing by language," in *Advances in Neural Information Processing Systems* (Long Beach, CA), 6594–6604.

Deco, G., and Rolls, E. T. (2004). A neurodynamical cortical model of visual attention and invariant object recognition. *Vis. Res.* 44, 621–642. doi: 10.1016/j.visres.2003.09.037

Deco, G., and Rolls, E. T. (2005). Neurodynamics of biased competition and cooperation for attention: a model with spiking neurons. *J. Neurophysiol.* 94, 295–313. doi: 10.1152/jn.01095.2004

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 353, 1245–1255. doi: 10.1098/rstb.1998.0280

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *arXiv [preprint]* arXiv:1810.04805.

Diamond, D. M. (2005). Cognitive, endocrine and mechanistic perspectives on non-linear relationships between arousal and brain function. *Nonlinearity Biolo Toxicol Med.* 3, 1–7. doi: 10.2201/nonlin.003.01.001

Driver, J. (2001). A selective review of selective attention research from the past century. *Br. J. Psychol.* 92, 53–78. doi: 10.1348/000712601162103

Elsayed, G., Kornblith, S., and Le, Q. V. (2019). "Saccader: improving accuracy of hard attention models for vision," in *Advances in Neural Information Processing Systems* (Vancouver, BC), 700–712.

Firat, O., Cho, K., and Bengio, Y. (2016). Multi-way, multilingual neural machine translation with a shared attention mechanism. *arXiv preprint arXiv:1601.01073*. doi: 10.18653/v1/N16-1101

Fong, R. C., Scheirer, W. J., and Cox, D. D. (2018). Using human brain activity to guide machine learning. *Sci. Rep.* 8:5397. doi: 10.1038/s41598-018-23618-6

Fries, P., Womelsdorf, T., Oostenveld, R., and Desimone, R. (2008). The effects of visual stimulation and selective visual attention on rhythmic neuronal synchronization in macaque area v4. *J. Neurosci.* 28, 4823–4835. doi: 10.1523/JNEUROSCI.4499-07.2008

Frings, C., Schneider, K. K., and Fox, E. (2015). The negative priming paradigm: an update and implications for selective attention. *Psychon. Bull. Rev.* 22, 1577–1597. doi: 10.3758/s13423-015-0841-4

Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (2007). Auditory attention–focusing the searchlight on sound. *Curr. Opin. Neurobiol.* 17, 437–455. doi: 10.1016/j.conb.2007.07.011

Fu, J., Zheng, H., and Mei, T. (2017). "Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI), 4438–4446.

Galassi, A., Lippi, M., and Torroni, P. (2019). Attention, please! a critical review of neural attention models in natural language processing. *arXiv [preprint]*. arXiv:1902.02181.

Gardiner, J. M., and Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Mem. Cogn.* 18, 579–583.

Gopher, D. (1996). Attention control: explorations of the work of an executive controller. *Cogn. Brain Res.* 5, 23–38.

Gottfried, J. A., and Dolan, R. J. (2003). The nose smells what the eye sees: crossmodal visual facilitation of human olfactory perception. *Neuron* 39, 375–386. doi: 10.1016/S0896-6273(03)00392-1

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., and Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends Cogn. Sci.* 17, 585–593. doi: 10.1016/j.tics.2013.09.001

Graves, A., Wayne, G., and Danihelka, I. (2014). Neural turing machines. *arXiv [preprint]*. arXiv:1410.5401.

Hayden, B. Y., and Gallant, J. L. (2009). Combined effects of spatial and feature-based attention on responses of v4 neurons. *Vis. Res.* 49, 1182–1187. doi: 10.1016/j.visres.2008.06.011

Hayhoe, M., and Ballard, D. (2005). Eye movements in natural behavior. *Trends Cogn. Sci.* 9, 188–194. doi: 10.1016/j.tics.2005.02.009

Heinke, D., and Humphreys, G. W. (2003). Attention, spatial representation, and visual neglect: simulating emergent attention and spatial memory in the selective attention for identification model (SAIM). *Psychol. Rev.* 110, 29–87. doi: 10.1037/0033-295X.110.1.29

Heinke, D., and Humphreys, G. W. (2005). Computational models of visual selective attention: a review. *Connect. Models Cogn. Psychol.* 1, 273–312. doi: 10.4324/9780203647110

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J.-H., and Welsh, T. N. (2019). No one knows what attention is. *Attent. Percept. Psychophys.* 81, 2288–2303. doi: 10.3758/s13414-019-01846-w

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 7132–7141.

Hutmacher, F. (2019). Why is there so much more research on vision than on any other sensory modality? *Front. Psychol.* 10:2246. doi: 10.3389/fpsyg.2019.02246

Itti, L., and Baldi, P. F. (2006). "Bayesian surprise attracts human attention," in *Advances in Neural Information Processing Systems* (Vancouver, BC), 547–554.

Itti, L., and Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci.* 2, 194–203. doi: 10.1038/35058500

Jaegle, A., Mehrpour, V., and Rust, N. (2019). Visual novelty, curiosity, and intrinsic reward in machine learning and the brain. *Curr. Opin. Neurobiol.* 58, 167–174. doi: 10.1016/j.conb.2019.08.004

Jain, S., and Wallace, B. C. (2019). Attention is not explanation. *arXiv [preprint]*. arXiv:1902.10186.

Johansen-Berg, H., and Lloyd, D. M. (2000). The physiology and psychology of selective attention to touch. *Front. Biosci.* 5, D894–D904. doi: 10.2741/A558

Johnson, M. H. (2001). Functional brain development in humans. *Nat. Rev. Neurosci.* 2, 475–483. doi: 10.1038/35081509

Kanwisher, N., and Wojciulik, E. (2000). Visual attention: insights from brain imaging. *Nat. Rev. Neurosci.* 1, 91–100. doi: 10.1038/35039043

Kaplan, F., and Hafner, V. V. (2006). The challenges of joint attention. *Interact. Stud.* 7, 135–169. doi: 10.1075/is.7.2.04kap

Keller, A. (2011). Attention and olfactory consciousness. *Front. Psychol.* 2:380. doi: 10.3389/fpsyg.2011.00380

Kelley, T. A., and Yantis, S. (2009). Learning to attend: effects of practice on information selection. *J. Vis.* 9:16. doi: 10.1167/9.7.16

Kelley, T. A., and Yantis, S. (2010). Neural correlates of learning to attend. *Front. Hum. Neurosci.* 4:216. doi: 10.3389/fnhum.2010.00216

Klein, J. T., Shepherd, S. V., and Platt, M. L. (2009). Social attention and the brain. *Curr. Biol.* 19, R958–R962. doi: 10.1016/j.cub.2009.08.010

Krauzlis, R. J., Lovejoy, L. P., and Zénon, A. (2013). Superior colliculus and visual spatial attention. *Annu. Rev. Neurosci.* 36, 165–182. doi: 10.1146/annurev-neuro-062012-170249

Lamme, V. A., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579. doi: 10.1016/S0166-2236(00)01657-X

Lee, S.-H., Shin, J.-K., and Lee, M. (2004). "Non-uniform image compression using biologically motivated saliency map model," in *Proceedings of the 2004 Intelligent Sensors, Sensor Networks and Information Processing Conference, 2004* (Melbourne, VIC), 525–530.

Lee, T. S., and Mumford, D. (2003). Hierarchical bayesian inference in the visual cortex. *JOSA A* 20, 1434–1448. doi: 10.1364/JOSAA.20.001434

Levi, H., and Ullman, S. (2020). Multi-task learning by a top-down control network. *arXiv [Preprint]*. arXiv:2002.03335.

Lin, M., Fu, J., and Bengio, Y. (2019). Conditional computation for continual learning. *arXiv [preprint]*. arXiv:1906.06635.

Lindsay, G. (2020). Convolutional neural networks as a model of the visual system: past, present, and future. *J. Cogn. Neurosci.* doi: 10.1162/jocn_a_01544. [Epub ahead of print].

Lindsay, G. W., and Miller, K. D. (2018). How biological attention mechanisms improve task performance in a large-scale visual system model. *eLife* 7:e38105. doi: 10.7554/eLife.38105

Lindsay, G. W., Rubin, D. B., and Miller, K. D. (2019). A simple circuit model of visual cortex explains neural and behavioral aspects of attention. *bioRxiv. [preprint]*. doi: 10.1101/2019.12.13.875534

Linsley, D., Shiebler, D., Eberhardt, S., and Serre, T. (2018). Learning what and where to attend. *arXiv [preprint]*. arXiv:1805.08819.

Liu, T., Slotnick, S. D., Serences, J. T., and Yantis, S. (2003). Cortical mechanisms of feature-based attentional control. *Cereb. Cortex* 13, 1334–1343. doi: 10.1093/cercor/bhg080

Lozito, J. P., and Mulligan, N. W. (2006). Exploring the role of attention during memory retrieval: effects of semantic encoding and divided attention. *Mem. Cogn.* 34, 986–998. doi: 10.3758/BF03193246

Luck, S. J., Chelazzi, L., Hillyard, S. A., and Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* 77, 24–42. doi: 10.1152/jn.1997.77.1.24

Luo, X., Roads, B. D., and Love, B. C. (2020). The costs and benefits of goal-directed attention in deep convolutional neural networks. *arXiv [preprint]*. arXiv:2002.02342.

Luong, M.-T., Pham, H., and Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. *arXiv preprint]*. arXiv:1508.04025.

Makeig, S., Jung, T.-P., and Sejnowski, T. J. (2000). Awareness during drowsiness: dynamics and electrophysiological correlates. *Can. J. Exp. Psychol.* 54, 266–273. doi: 10.1037/h0087346

Mancas, M., Ferrera, V. P., Riche, N., and Taylor, J. G. (2016). *From Human Attention to Computational Attention, Vol. 2*. New York, NY: Springer.

Maninis, K.-K., Radosavovic, I., and Kokkinos, I. (2019). "Attentive single-tasking of multiple tasks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Long Beach, CA), 1851–1860. doi: 10.1109/CVPR.2019.00195

Manohar, S. G., Zokaei, N., Fallon, S. J., Vogels, T., and Husain, M. (2019). Neural mechanisms of attending to items in working memory. *Neurosci. Biobehav. Rev.* 101, 1–12. doi: 10.1016/j.neubiorev.2019.03.017

Marks, L. E., and Wheeler, M. E. (1998). Attention and the detectability of weak taste stimuli. *Chem. Senses* 23, 19–29. doi: 10.1093/chemse/23.1.19

Maunsell, J. H. (2015). Neuronal mechanisms of visual attention. *Annu. Rev. Vis. Sci.* 1, 373–391. doi: 10.1146/annurev-vision-082114-035431

Miller, E. K., and Buschman, T. J. (2014). "Neural mechanisms for the executive control of attention," in *The Oxford Handbook of Attention*, eds A. C. Nobre and S. Kastner (Oxford, UK: Oxford University Press).

Mirza, M. B., Adams, R. A., Friston, K., and Parr, T. (2019). Introducing a bayesian model of selective attention based on active inference. *Sci. Rep.* 9:13915. doi: 10.1038/s41598-019-50138-8

Mitchell, J. F., Sundberg, K. A., and Reynolds, J. H. (2007). Differential attention-dependent response modulation across cell classes in macaque visual area v4. *Neuron* 55, 131–141. doi: 10.1016/j.neuron.2007.06.018

Mnih, V., Heess, N., Graves, A., et al. (2014). "Recurrent models of visual attention," in *Advances in Neural Information Processing Systems* (Montreal, QC), 2204–2212.

Moore, T., Armstrong, K. M., and Fallah, M. (2003). Visuomotor origins of covert spatial attention. *Neuron* 40, 671–683. doi: 10.1016/S0896-6273(03)00716-5

Murray, M. M., De Santis, L., Thut, G., and Wylie, G. R. (2009). The costs of crossing paths and switching tasks between audition and vision. *Brain Cogn.* 69, 47–55. doi: 10.1016/j.bandc.2008.05.004

Muzzio, I. A., Kentros, C., and Kandel, E. (2009). What is remembered? Role of attention on the encoding and retrieval of hippocampal representations. *J. Physiol.* 587, 2837–2854. doi: 10.1113/jphysiol.2009.172445

Myachykov, A., and Posner, M. I. (2005). "Attention in language," in *Neurobiology of Attention*, eds L. Itti, G. Rees, and J. K. Tsotsos (Burlington, MA: Elsevier), 324–329.

Noudoost, B., Chang, M. H., Steinmetz, N. A., and Moore, T. (2010). Top-down control of visual attention. *Curr. Opin. Neurobiol.* 20, 183–190. doi: 10.1016/j.conb.2010.02.003

O'Craven, K. M., Downing, P. E., and Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature* 401, 584–587. doi: 10.1038/44134

Oken, B. S., Salinsky, M. C., and Elsas, S. (2006). Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clin. Neurophysiol.* 117, 1885–1901. doi: 10.1016/j.clinph.2006.01.017

Olivers, C. N., and Eimer, M. (2011). On the difference between working memory and attentional set. *Neuropsychologia* 49, 1553–1558. doi: 10.1016/j.neuropsychologia.2010.11.033

Paneri, S., and Gregoriou, G. G. (2017). Top-down control of visual attention by the prefrontal cortex. Functional specialization and long-range interactions. *Front. Neurosci.* 11:545. doi: 10.3389/fnins.2017.00545

Pashler, H. (2000). "Task switching and multitask performance," in *Control of Cognitive Processes: Attention and Performance XVIII*, eds S. Monsell and J. Driver (MIT Press), 277. doi: 10.1002/acp.849

Pestilli, F., Viera, G., and Carrasco, M. (2007). How do attention and adaptation affect contrast sensitivity? *J. Vis.* 7, 9.1–9.12. doi: 10.1167/7.7.9

Phaf, R. H., Van der Heijden, A., and Hudson, P. T. (1990). SLAM: a connectionist model for attention in visual selection tasks. *Cogn. Psychol.* 22, 273–341. doi: 10.1016/0010-0285(90)90006-P

Poldrack, R. A., Sabb, F. W., Foerde, K., Tom, S. M., Asarnow, R. F., Bookheimer, S. Y., et al. (2005). The neural correlates of motor skill automaticity. *J. Neurosci.* 25, 5356–5364. doi: 10.1523/JNEUROSCI.3880-04.2005

Posner, M. I. (2008). Measuring alertness. *Ann. N. Y. Acad. Sci.* 1129, 193–199. doi: 10.1196/annals.1417.011

Radulescu, A., Niv, Y., and Ballard, I. (2019). Holistic reinforcement learning: the role of structure and attention. *Trends Cogn. Sci.* 23, 278–292. doi: 10.1016/j.tics.2019.01.010

Rao, R. P. (2005). Bayesian inference and attentional modulation in the visual cortex. *Neuroreport* 16, 1843–1848. doi: 10.1097/01.wnr.0000183900.92901.fc

Rebuffi, S.-A., Bilen, H., and Vedaldi, A. (2017). "Learning multiple visual domains with residual adapters," in *Advances in Neural Information Processing Systems* (Long Beach, CA), 506–516.

Ren, M., and Zemel, R. S. (2017). "End-to-end instance segmentation with recurrent attention," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Long Beach, CA), 6656–6664.

Reynolds, G. D., and Romano, A. C. (2016). The development of attention systems and working memory in infancy. *Front. Syst. Neurosci.* 10:15. doi: 10.3389/fnsys.2016.00015

Reynolds, J. H., and Heeger, D. J. (2009). The normalization model of attention. *Neuron* 61, 168–185. doi: 10.1016/j.neuron.2009.01.002

Rizzolatti, G., Riggio, L., Dascola, I., and Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. *Neuropsychologia* 25, 31–40. doi: 10.1016/0028-3932(87)90041-8

Roelfsema, P. R., and Houtkamp, R. (2011). Incremental grouping of image elements in vision. *Attent. Percept. Psychophys.* 73, 2542–2572. doi: 10.3758/s13414-011-0200-0

Roelfsema, P. R., Lamme, V. A., and Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature* 395, 376–381. doi: 10.1038/26475

Rossi, A. F., and Paradiso, M. A. (1995). Feature-specific effects of selective visual attention. *Vis. Res.* 35, 621–634. doi: 10.1016/0042-6989(94)00156-G

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* 115, 211–252. doi: 10.1007/s11263-015-0816-y

Saalmann, Y. B., Pinsk, M. A., Wang, L., Li, X., and Kastner, S. (2012). The pulvinar regulates information transmission between cortical areas based on attention demands. *Science* 337, 753–756. doi: 10.1126/science.1223082

Saenz, M., Buracas, G. T., and Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nat. Neurosci.* 5, 631–632. doi: 10.1038/nn876

Sajedin, A., Menhaj, M. B., Vahabie, A.-H., Panzeri, S., and Esteky, H. (2019). Cholinergic modulation promotes attentional modulation in primary visual cortex-a modeling study. *Sci. Rep.* 9:20186. doi: 10.1038/s41598-019-56608-3

Samuels, E. R., and Szabadi, E. (2008). Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part i: principles of functional organisation. *Curr. Neuropharmacol.* 6, 235–253. doi: 10.2174/157015908785777229

Schweisfurth, M. A., Schweizer, R., and Treue, S. (2014). Feature-based attentional modulation of orientation perception in somatosensation. *Front. Hum. Neurosci.* 8:519. doi: 10.3389/fnhum.2014.00519

Shapiro, K. L., Raymond, J., and Arnell, K. (1997). The attentional blink. *Trends Cogn. Sci.* 1, 291–296. doi: 10.1016/S1364-6613(97)01094-2

Sharma, S., Kiros, R., and Salakhutdinov, R. (2015). Action recognition using visual attention. *arXiv [preprint]*. arXiv:1511.04119.

Shenhav, A., Botvinick, M. M., and Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* 79, 217–240. doi: 10.1016/j.neuron.2013.07.007

Shipp, S. (2004). The brain circuitry of attention. *Trends Cogn. Sci.* 8, 223–230. doi: 10.1016/j.tics.2004.03.004

Soto, D., Heinke, D., Humphreys, G. W., and Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 248–261. doi: 10.1037/0096-1523.31.2.248

Soto, D., Hodsoll, J., Rotshtein, P., and Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends Cogn. Sci.* 12, 342–348. doi: 10.1016/j.tics.2008.05.007

Spence, C. (2009). Explaining the colavita visual dominance effect. *Prog. Brain Res.* 176, 245–258. doi: 10.1016/S0079-6123(09)17615-X

Spence, C., and Driver, J. (2004). *Crossmodal Space and Crossmodal Attention*. Oxford, UK: Oxford University Press.

Sreenivasan, V., and Sridharan, D. (2019). Subcortical connectivity correlates selectively with attention's effects on spatial choice bias. *Proc. Natl. Acad. Sci. U.S.A.* 116, 19711–19716. doi: 10.1073/pnas.1902704116

Stollenga, M. F., Masci, J., Gomez, F., and Schmidhuber, J. (2014). "Deep networks with internal selective attention through feedback connections," in *Advances in Neural Information Processing Systems* (Montreal, QC), 3545–3553.

Strezoski, G., van Noord, N., and Worring, M. (2019). Many task learning with task routing. *arXiv preprint arXiv:1903.12117*. doi: 10.1109/ICCV.2019.00146

Tatler, B. W., Baddeley, R. J., and Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vis. Res.* 45, 643–659. doi: 10.1016/j.visres.2004.09.017

Toneva, M., and Wehbe, L. (2019). "Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain)," in *Advances in Neural Information Processing Systems*, 14928–14938.

Treisman, A., Vieira, A., and Hayes, A. (1992). Automaticity and preattentive processing. *Am. J. Psychol.* 105, 341–362. doi: 10.2307/1423032

Treue, S., and Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399:575. doi: 10.1038/21176

Turk-Browne, N. B., Jungé, J. A., and Scholl, B. J. (2005). The automaticity of visual statistical learning. *J. Exp. Psychol.* 134, 552–564. doi: 10.1037/0096-3445.134.4.552

Uncapher, M. R., and Rugg, M. D. (2005). Effects of divided attention on fmri correlates of memory encoding. *J. Cogn. Neurosci.* 17, 1923–1935. doi: 10.1162/089892905775008616

van Zoest, W., and Donk, M. (2005). The effects of salience on saccadic target selection. *Vis. Cogn.* 12, 353–375. doi: 10.1080/13506280444000229

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). "Attention is all you need," in *Advances in Neural Information Processing Systems* (Long Beach, CA), 5998–6008.

Wagner, A. D., Shannon, B. J., Kahn, I., and Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends Cogn. Sci.* 9, 445–453. doi: 10.1016/j.tics.2005.07.001

Wang, Q., Zhang, J., Song, S., and Zhang, Z. (2014). "Attentional neural network: Feature selection using cognitive feedback," in *Advances in Neural Information Processing Systems* (Montreal, QC), 2033–2041.

Watanabe, W. (2003). Is subliminal learning really passive? *Nature* 422:36. doi: 10.1038/422036a

Wiegreffe, S., and Pinter, Y. (2019). Attention is not not explanation. *arXiv [preprint]*. arXiv:1908.04626.

Wolf, L., Guttmann, M., and Cohen-Or, D. (2007). "Non-homogeneous content-driven video-retargeting," in *2007 IEEE 11th International Conference on Computer Vision* (Rio de Janeiro), 1–6.

Wolfe, J. M., and Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.* 5, 495–501. doi: 10.1038/nrn1411

Wood, S., Sage, J. R., Shuman, T., and Anagnostaras, S. G. (2014). Psychostimulants and cognition: a continuum of behavioral and cognitive activation. *Pharmacol. Rev.* 66, 193–221. doi: 10.1124/pr.112.007054

Wyatte, D., Jilk, D. J., and O'Reilly, R. C. (2014). Early recurrent feedback facilitates visual object recognition under challenging conditions. *Front. Psychol.* 5:674. doi: 10.3389/fpsyg.2014.00674

Xu, H., and Saenko, K. (2016). "Ask, attend and answer: exploring question-guided spatial attention for visual question answering," in *European Conference on Computer Vision* (Amsterdam: Springer), 451–466.

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., et al. (2015). "Show, attend and tell: Neural image caption generation with visual attention," in *International Conference on Machine Learning* (Lille), 2048–2057.

Yang, Z., He, X., Gao, J., Deng, L., and Smola, A. (2016). "Stacked attention networks for image question answering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV), 21–29.

Zagoruyko, S., and Komodakis, N. (2016). Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer. *arXiv [preprint]*. arXiv:1612.03928.

Zelinsky, G. J., Chen, Y., Ahn, S., Adeli, H., Yang, Z., Huang, L., et al. (2020). Predicting goal-directed attention control using inverse-reinforcement learning. *arXiv [preprint]*. arXiv:2001.11921.

Zhao, X., Li, H., Shen, X., Liang, X., and Wu, Y. (2018). "A modulation module for multi-task learning with applications in image retrieval," in *Proceedings of the European Conference on Computer Vision (ECCV)* (Munich), 401–416.

Zhou, H., and Desimone, R. (2011). Feature-based attention in the frontal eye field and area V4 during visual search. *Neuron* 70, 1205–1217. doi: 10.1016/j.neuron.2011.04.032

Zhou, H., Schafer, R. J., and Desimone, R. (2016). Pulvinar-cortex interactions in vision and attention. *Neuron* 89, 209–220. doi: 10.1016/j.neuron.2015.11.034

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MR declared a past co-authorship with the author GL to the handling Editor.

title: Attention in Psychology, Neuroscience, and Machine Learning author: Grace W. Lindsay year: 2020 journal: Frontiers in Computational Neuroscience month: April volume: 14 article: 29

- 2019JCSS

心理学，神経科学，機械学習における注意

Grace W. Lindsay

要約

注意は，限られた計算資源を柔軟にコントロールする重要な能力である。神経科学や心理学の分野では，意識，警戒，注意，実行制御，学習など，他の多くのトピックと関連して研究されている。また，最近では，機械学習のいくつかの領域でも応用されている。生物学的注意研究と，人工ニューラルネットワークを強化するツールとしての注意の使用との関係は，必ずしも明確ではない。このレビューでは，まず，神経科学や心理学の文献で注意がどのように概念化されているかを概観する。次に，機械学習における注意の使用例をいくつか取り上げ，生物学的に対応するものがある場合はそれを示す。最後に，複雑で統合的なシステムを構築するために，人工的な注意力を生物学からさらに引き出す方法について検討する。

1. はじめに

注意は，公に広く議論され，科学的にも広く研究されているトピックである。心理学，神経科学，そして最近では機械学習を含む複数の分野で，様々な定義がなされている (Chun et al.2011; Cho et al. 2015)。ウィリアム・ジェームズは実験心理学の黎明期にこう書いた:「注意とは何かは，誰もが知っている。注意とは，同時に可能と思われる複数の対象物や思考の流れのうち，1つを明確かつ鮮明な形で心に刻むことである」。ジェームズがこの言葉を残して以来，このプロセスをより正確に定義し，定量化するために多くの試みがなされており，また，このプロセスを生み出す根本的な精神的・神経的構造も明らかにされている。しかし，ひとつの概念として語られるものを研究するために，さまざまな実験的アプローチや概念が氾濫していることが，研究者の反発を招いている。概念に対するより進化に基づいたアプローチを主張する最近の論文のタイトルにもあったように，「注意とは何かは誰にもわからない」のである（Hommel et al. 2019)

注意は、明確で統一された概念とは程遠いものである。しかし，多くの曖昧な定義があり，時には相反する定義があるにもかかわらず，脳や，最近では AI システムの情報処理にとって非常に重要であることが明らかになっている，注意の中核となる性質がある。注意とは，限られた計算機資源を柔軟に制御することである。なぜ注意の資源が限られているのか，どのように制御するのがベストなのかは，ユースケースによって異なる。しかし，情報の流れを動的に変化させたり，ルーティングしたりする能力は，あらゆるシステムの適応性にとって明らかに有益である。

注意が脳内で多くの役割を果たしていることを考えれば，人工ニューラルネットワークに注意が加わるのは当然のことと言える。人工ニューラルネットワークは，ニューロンの基本的な入出力機能を模倣して設計された，個々のユニットからなる並列処理システムであり，現在，機械学習や人工知能 (AI) の分野で主流となっているモデルである。当初は注目されずに構築されていたが，現在ではネットワークの表現や構造を動的に再構成するためのさまざまな仕組みが追加されている。

続く 第 2 節では，神経科学と心理学における「注意」という言葉のさまざまな使用法と，他の一般的な神経科学のトピックとの関連性について概説する。全体を通して，限られた資源をコントロールする方法としての注意の概念が強調される。行動学的研究は，注意の能力と限界を示すために使用され，神経メカニズムは，これらの行動的効果が現れる物理的手段を示す。第 3 節では，機械学習における注意の研究状況を要約し，AI の注意と生物学的な注意との関係がある場合にはそれを示す。第 4 節では，生物学的注意から得られた知見が人工的な注意に影響を与える追加的な方法を紹介する。

本概説論文の第一の目的は，AI や機械学習の分野の研究者に，神経科学や心理学において注意がどのように概念化され，研究されているかを理解してもらい，実りあるインスピレーションを得られるようにすることである。第二の目的は，生物学的注意を研究している研究者に，これらのプロセスが AI システムでどのように運用されているかを伝えることである。これにより，生物学的知見の機能的意味合いについての考え方に影響を与える可能性があるだろう。

2. 神経科学と心理学における注意

注意の科学的研究は心理学から始まった。慎重に行動実験を行うことで，さまざまな状況下での注意の傾向や能力を正確に示すことができる。認知科学や認知心理学では，こうした観察結果をもとに，どのような心のプロセスがそのような行動パターンを生み出すのかをモデル化することを目指している。これまでに，さまざまな基礎的メカニズムを想定した多くの単語モデルや計算モデルが作られてきた (Driver, 2001; Borji and Itti, 2012)。

人間以外の霊長類における単一細胞の神経生理学の影響や，EEG，fMRI，MEG などの非侵襲的な脳活動のモニタリング手段により，基礎となる神経プロセスの直接観察が可能になった。これにより，注意に関連した神経反応の特定の特徴を再現できる神経回路の計算モデルが構築された (Shipp, 2004)。

以下では，注意に関するいくつかの異なるクラスについて，行動的および神経的な知見を説明する。

2.1. 覚醒，アラートネス，警戒，としての注意

最も一般的な形では， 注意とはすなわち， 周囲の環境に関与するための全体的な警戒心や能力のレベルであると言える。 このようにして， 注意は覚醒や睡眠・覚醒スペクトルと相互作用する。 心理学における「Vigilance 警戒」は、注意を持続する能力を意味しており、これも関連している。 なお，これらの言葉は同じ意味で使われているが， 異なるニッチな文献ではより具体的に使われていることもありる (Oken et al, 2006)。

睡眠覚醒サイクルのさまざまな段階， 睡眠不足， 鎮静剤の服用中の被験者を研究することで， このような注意の形態がどのように変化し， その結果どのような行動が起こるのかを知ることができる。 スクリーン上の特定の領域にボールを置くなど， 持続的な注意力を必要とする反復的な課題を被験者に与えることで， 研究者たちは， 脳波信号の変化と関連して， 眠気を感じている患者の成績が長時間低下することを観察した（Makeig et al, 2000）。 しかし， 課題をより魅力的にすることで， 眠気や鎮静状態でも成績を向上させることができる方法がある。 課題を実行することで得られる報酬を増やす， 目新しさや不規則性を加える， ストレスを導入するなどである (Okenet al., 2006)。 したがって， 一般的な注意には限られた蓄えがあり， 平凡な課題や十分な報酬が得られない課題の場合には展開されないが， より有望な課題や興味深い課題の場合には呼び出せるようである。

興味深いことに、覚醒度が高ければ良いというわけではない。 Yerkes-Dodson曲線（図1B）は逆U字型で、十分に難易度の高い課題に対する覚醒度の関数としてパフォーマンスを表している。 覚醒度が低いとパフォーマンスは低下し， 中程度になると良好になり， 高くなると再び低下する。 オリジナルの研究では， マウスに電気ショックを与えて覚醒度を変化させたが， この結果は他の測定方法でも繰り返されている (Diamond, 2005)。 アデロールやカフェインなどの精神刺激剤が， ある人にはある用量で集中力を高める働きをするが， 他の人には有害になる理由を説明できるかもしれない (Woodet al., 2014)。

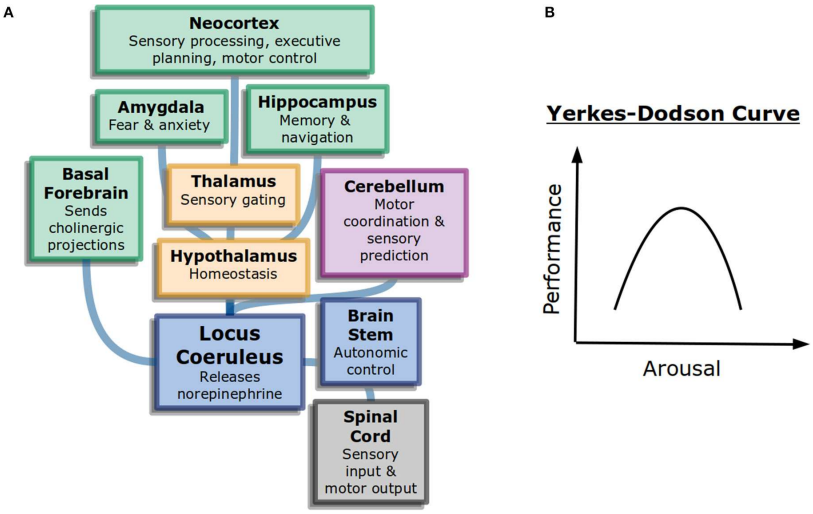

図 1 青斑核の細胞は， ノルエピネフリン (ノルアドレナリンとしても知られている) を， 他の神経調節系を含む， さまざまな機能を持つ脳の多くの部分に放出する。 これが全体的な覚醒に寄与する (Samuels and Szabadi, 2008)。 ここでの色は， 前脳 (緑)， 間脳(黄)， 脳幹(青) といった脳の各部位を表している。 (B）ヤーケス・ドッドソン曲線は，覚醒度と課題遂行能力の非線形関係を表している。

睡眠・覚醒サイクルの基盤となる神経回路は， 主に脳幹にある (Coenen, 1998)。 これらの回路は， 視床と先天性皮質への情報の流れを制御する。 さらに， 神経調節系は全般的な注意力の制御に大きな役割を果たしている。 ノルエピネフリン， アセチルコリン， ドーパミンは， それぞれ， 警戒心， 重要な情報への方向付け， 注意の実行制御に影響を与えると考えられている (Posner, 2008)。 神経調節物質の解剖学的構造は， その機能とも一致している。 例えば， ノルエピネフリンを放出する神経細胞は， 細胞体は脳幹にあるが， 脳全体に非常に広く分布しており， 情報処理を広範囲に制御することができる (図1A)。

2.2. 感覚的注意

全体的な覚醒度に加えて， 覚醒している被験者は特定の感覚入力に対して選択的に注意を展開することができる。 特定の感覚システムの中で注意を研究することで， 刺激と注意の位置の両方を厳密に制御することができる。 一般的に， このタイプの注意を調べるには， 使用する課題がかなり難しいものである必要がある。 例えば， 変化検出課題では， 2 つの刺激の間の検出されるべき違いは非常に小さいかもしれない。 より一般的には， 課題の難しさは， 刺激を短時間提示するか， 弱く提示することで達成できる。

システム神経科学や心理学における注意研究の大部分は， 特に視覚的注意を中心としている (Kanwisher and Wojciulik, 2000)。 これらの分野では， 他の感覚システムよりも視覚処理の研究を重視するという一般的な傾向 (Hutmacher, 2019) と， 霊長類の脳において視覚が支配的な役割を果たしていることを反映していると考えられる。 さらに， 注意のより一般的で認知的な側面を扱う研究でも， 視覚刺激が頻繁に使用されている。

視覚的注意は， 空間的注意と特徴に基づく注意に大別される。

2.2.1. 視覚的空間的注意

サッカードとは 1 秒間に数回行われる小さくて速い眼球運動のことである。 眼窩は網膜上で最も高い解像度を持つため， 眼窩をどこに配置するかは，限られた計算機資源をどこに配置するかという選択に他ならない。 このように， 眼球運動は注意の所在を示す。 注意の移動が外に向かって見えることから「顕在的視覚注意」と呼ばれている。

研究者たちは， 異なる画像を提示したときの眼球運動を追跡することで， 自動的に注意を引きつける画像パターンを特定してきた。 このようなパターンは， エッジの向き， 空間周波数， 色の対比， 強度， 動きなどによって定義される (Itti and Koch, 2001)。 注意を引く画像領域は「顕著」とみなされ， 「ボトムアップ」方式で計算される。 つまり， 意識的な処理や努力を必要とせず， 視覚システムに組み込まれた特徴検出器の結果であると考えら

れる。 そのため， 顕著性 (顕在性) は非常に高速に計算される。 さらに， どの領域が顕著なのか， 特に最初の数回のサッケードで識別された領域については， 異なる被験者でも一致する傾向がある（Tatler et al, 2005)。

顕著な領域は， 画像をどのように見るかという特定の指示を与えられていない「自由観察」の状況で調べることができる。 特定の課題が与えられると， ボトムアップ注意とトップダウン注意の相互作用が明らかになる。 例えば， 配列の中の特定の視覚目標にサッケードするように指示された場合，被験者は，代わりに特に顕著な妨害刺激に誤ってサッケードすることがある (van Zoest and Donk, 2005)。 より一般的には， 複雑な自然画像を見ているときに， 高レベルの課題 (例えば， 人物の年齢を評価したり， 社会経済的地位を推測したりするような課題) を与えられた場合， 課題の指示がサッケードのパターンに大きな影響を与える可能性がある。 さらに， 被験者がサンドイッチを作るような実世界の課題を実行する際の自然な視線移動のパターンは， 基本的な認知プロセスに関する洞察を与えてくれる (Hayhoe and Ballard, 2005)。

被験者は， 複数回のサッケードを連続して行う必要がある場合， 最近見た場所には戻らない傾向があり， その場所で何か関連することが起こっても反応が鈍くなることがある。 この現象は「戻り抑制」として知られている (Itti and Koch, 2001)。 このような行動は， 視覚システムに，元々最も顕著であると考えられた画像領域を利用するだけでなく， 他の領域も探索するように促す。 また， サッケードを生成するシステムには， 一種の記憶が必要である。 これは， 最近見た場所の表現を短期的に抑制することで実現されると考えられている。

眼球運動は視覚的注意を制御する有効な手段であるが， 唯一の選択肢ではない。 「隠れた」空間的注意とは， 焦点の位置を明らかに変化させることなく， 異なる空間的位置の処理を強調する方法である。 一般に， 隠れた空間注意の研究では， 被験者は課題中， 中心点を固定しなければならない。 このとき， 被験者は， 自分の視覚課題に関連した刺激が現れる可能性の高い周辺視野の場所に隠密に注意を払うように促される。 例えば， 方位識別課題では， 空間的な手がかりが与えられた後， 手がかりとなった場所に方位のある格子が点滅し， 被験者はその方位を示す必要がある。 無効手がかり試行 (手がかりのない場所に刺激が現れる場合) では， 被験者は有効手がかり試行 (または手がかりのない試行) よりも成績が悪くなる (Anton-Erxleben and Carrasco, 2013)。 これは， 隠れた空間的注意が， 柔軟に展開できる限られた資源であり， 視覚情報の処理を助けることを示している。

隠れた空間的注意は， 特定の領域が他の領域を犠牲にしてさらなる処理のために選択されるという意味で， 選択的である。 これは， 注意の「スポットライト」と呼ばれている。 重要なのは， 過剰な注意とは対照的， 隠れた注意では， 視覚システムへの入力は同じでも， その入力の処理は柔軟に選択されるということである。

隠れた空間的注意は， ボトムアップの顕著性によっても影響を受ける。 無関係だが顕著な物体が， 課題に関連する刺激がある場所で点滅した場合， 無関係な刺激によって引き寄せられた外因性の空間的注意は， 課題に関連する刺激に適用され， 成績に利益をもたらす可能性がある。 しかし， 無関係な場所でフラッシュされた場合は， 助けにはならず， 成績に悪影響を与える可能性がある(Berger et al, 2005)。 ボトムアップ/外因性注意は時間経過が早く， 妨害刺激が現れてから 80〜130 msの間， 隠れた注意に影響を与える (Anton-Erxleben and Carrasco, 2013)。

注意理論の中には， 隠れた空間的注意は顕在的注意を導くために存在するというものがある。 特に， 運動前注意理論では， サッケードを計画する神経回路と隠蔽された空間的注意を制御する神経回路は同じであると考えられている (Rizzolatti et al, 1987)。 眼球運動の制御には， 前頭眼野 (FEF) が関与していることが知られている。 FEF のニューロンを， 眼球運動を誘発するには低すぎるレベルで刺激すると， 隠れた注意に似た効果が得られることが示されている (Moore et al, 2003)。 このように， 隠れた注意は， あからさまに見るべき場所を決めるための手段であると考えられる。 目の動きは， 秘密にしておいた方がよい知識や意図に関する情報を伝えるので， 隠密に注意を払う能力は， 社会的な種にも役立つかもしれない (Klein et al, 2009)。

隠れた空間注意の神経相関を研究するために， 研究者は， 注意の手がかりの違いだけに基づいて (刺激のボトムアップの特徴の違いではなく) 神経活動のどの側面が異なるかを特定する。 注意が記録されたニューロンの受容野に向かって手がかり付けされる試行では， 神経活動の多くの変化が観察されている (Noudoost et al, 2010; Maunsell, 2015)。 一般的に報告されている所見は， 発火率の増加で， 典型的には 20〜30％の増加である (Mitchell et al, 2007)。 しかし， 変化の正確な大きさは， 調査する皮質領域によって異なり， 後の領域ほど強い変化を示す (Luck et al, 1997; Noudoost et al, 2010)。 注意は， 神経発火のばらつきにも影響を与えることが知られている。 特に Fano Factor で測定される試行間の変動性を減少させ， ニューロン対間のノイズ相関を減少させる。 さらに， 注意はニューロンの電気生理学的特性にも影響を与え， バースト的に発火する可能性を減少させ， 個々の活動電位の高さを減少させることがわかっている (Anderson et al, 2013)。

一般に， 注意に伴う変化は， 注意した刺激を表すニューロンの信号対雑音比を増加させると考えられているが， 脳領域間のコミュニケーションにも影響を与える可能性がある。 この目的のためには， 注意が神経の同期に及ぼす影響が重要である。 視覚野では、注意によって、ガンマ帯（30〜70Hz）のスパイクのコヒーレンスが高まることが示されている（Fries et al, 2008)。 ニューロンのグループが同期して発火すると， 下流の共有領域に影響を与える能力が高まる。 さらに， 注意が領域間のコミュニケーションを直接調整している可能性もある。 2 つの視覚野の間の同期活動は， コミュニケーションが活発になっていることを示すことがあるす。 例えば， 注意は V1 野と V4 野の参加刺激を表すニューロン間の同期を高めることが示されている (Bosman et al. 2012)。 この領域間の同期の制御は， 視床枕によって行われているようだ (Saalmann et al, 2012)。

注意が視覚経路のニューロンにどのような影響を与えるかを調べるだけでなく， トップダウンの注意の源を探る研究も行われている (Noudoost et al, 2010; Miller and Buschman, 2014)。 ボトムアップの注意の処理は， 外側頭頂内野 (LIP) で生成される 顕著性地図に集約されるようである。 この領域の細胞は， 課題とは無関係だが重要な妨害物を含む， 重要な刺激が受容野にあるときに反応する。 一方， FEF のような前頭前野は， 空間的注意のトップダウン制御に必要な信号を格納しているようで， 妨害にはあまり反応しない。

感覚的注意の神経相関に関する研究の多くは大脳皮質を中心に行われているが， 皮質下の領域も注意制御や成績向上に強い役割を果たしているようである。 特に， 上丘は隠れた，および顕在的空間的注意を助け， この領域の不活性化は注意を損なう可能性がある(Krauzlis et al, 2013)。 また， 上述したように， 上丘は， 特に大脳皮質に対するゲーティング効果に関して， 注意に役割を果たしている (Zhou et al, 2016)。

2.2.2. 視覚的特徴への注意

特徴注意は， 隠れた選択的注意の一種である。 特徴注意では， 特定の場所に注意を向けるのではなく， 特定の色， 形， 方向 などの視覚的特徴に注意を向けるように試行ごとに指示される。 課題の目的は， 手がかりとなる特徴が画面上に存在するかどうかを検出したり， その特徴の別の性質を読み取ったりすることである (例えば「正方形の色は何か?」“ への答えは， まず正方形に注意が向くはずである)。 このように， 参加した特徴についての有効な手がかりは， 成績を向上させます。 例えば， ある特定の方向に注意を向けると， その方向の微弱なグレーティングを他の方向よりもよく検出することができた (Rossi and Paradiso, 1995)。 全体的な課題 (例: 方向性のある縞模様の検出) は変わらないが， 特定の指示 (90◦ の縞模様の検出 vs. 60◦ vs. 30◦) は， 個々の試行ごとに， あるいは場合によってはブロック毎にキューイングされる。 試行毎のキューに成功したことで， この形式の注意は速いタイムスケールで柔軟に展開できることがわかった。

視覚探索課題も，特徴に基づく注意を活性化すると考えられている (図2)。この課題では，スクリーン上に配列された刺激が表示され，被験者は手がかりとなる刺激の位置を目の動きで示す必要がある (頻繁に目を動かす)。被験者は通常、タスク中にサッケードをしながら手がかりとなる刺激を探すことができるので、このタスクは、隠れた特徴に基づく注意と顕在的な注意を組み合わせたものである。実際，サッケード選択に関与する領域であるFEFでは，トップダウンの特徴に基づく注意のシグナルが見つかっています (Zhou and Desimone, 2011)。例えば，複数の黒い図形の中に1つだけ赤い図形があるとすぐに注目されるように，ある種の特徴はポップアウト効果を生み出すことがあるため，視覚探索課題ではボトムアップ的な注意が働くことになり，課題によってはそれを抑制する必要がある (Wolfe and Horowitz, 2004)。

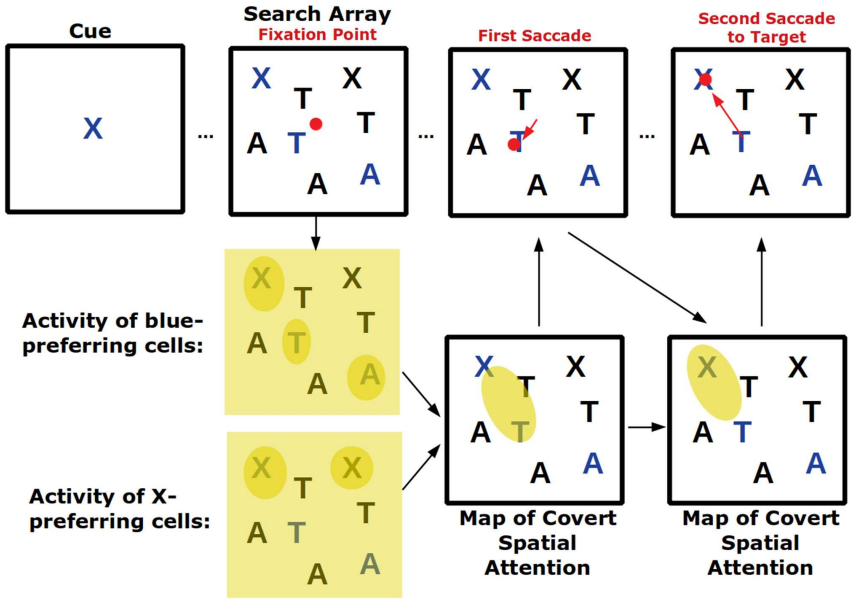

図 2 視覚探索課題では，さまざまな形で視覚的な注意が働く。上列は，視覚探索課題の進行を示している。まず，視覚探索の対象を示す手がかり (この場合青いX) が表示される。次に，多くの非標的を含む探索配列が現れる。青やXの形を表す細胞にトップダウンの特徴的な注意を向けると，視覚野全体で発火が増加するが，青やXが実際に存在する場所で発火が最も強くなる。これらの神経反応は，サッケード前に視覚空間を探索するために使用できる隠れた空間注意地図を生成する役割を果たす。最初のサッケードで顕在意識がシフトした後，隠れた意識マップが作り直される。最後に，ターゲットの位置を確認し，サッケードに成功する。もし，視覚的配列にポップアウトした刺激 (例えば緑のO) が含まれていたら，ボトムアップ方式で隠れた空間的注意を捉え，さらに誤ったサッカードを引き起こしたかもしれない。

視覚系における特徴に基づく注意の神経効果は，一般的に空間的注意のそれと似ている。例えば，注意された特徴を表すニューロンは発火率が増加し，非常に異なる特徴を表すニューロンは発火率が抑制される (Treue and Trujillo, 1999)。しかし，空間的注意とは対照的に，特徴に基づく注意は空間的に大域的である。つまり，特定の特徴に注意を向けると，視覚空間のあらゆる場所でその特徴を表すニューロンの活動が調整される (Saenz et al., 2002)。空間的注意と特徴的注意のもう一つの違いは，トップダウンの注意源が視覚系の正しいニューロンをどのように標的にするかという問題である。網膜トピック地図では，近くの細胞が近くの空間的な位置を表しているので，空間的なターゲティングは簡単だが，好ましい視覚的特徴に応じて細胞がきれいに整理されているわけではない。

空間的注意と特徴的注意の効果は相加的であると考えられる (Hayden and Gallant, 2009)。さらに，特徴注意も空間注意も，視覚野の分割正規化を実行する局所神経回路に作用することで効果を生み出すと考えられている (Reynolds and Heeger, 2009)。モデリングの研究では，トップダウン接続がこれらの回路の細胞に対象となるシナプス入力を提供すると仮定することで，選択的注意の神経効果の多くを捉えることができることが示されている（Lindsay et al, 2019）。しかし，神経調節物質であるアセチルコリンの効果に依存するモデルでも，注意の神経相関を再現することができる (Sajedin et al, 2019)。

トップダウンの特徴に基づく注意の潜在的な源は，持続的な活動が出席した特徴をエンコードする前頭前野に見出されている (Bichot et al, 2015; Paneri and Gregoriou, 2017)。腹部前頭葉領域を非活動化すると，探索課題の成績が低下する。前頭前野からは，上位の視覚野が下位の視覚野に入力を送るという逆階層的な方法で注意信号が伝わると考えられている (Ahissar and Hochstein, 2000)。

特徴注意と密接に関連する話題として，物体注意がある。この場合，注意は視覚刺激の前に抽象的な特徴に展開されるのではなく，視覚シーンの中の特定の物体に適用される (Chen, 2012)。視覚階層の活動の最初のフィードフォワード経路は，背景との明確で顕著な違いを持つ対象物があれば，事前に対象物を背景から分離して，視野内に並列に配置することができる。背景との違いが顕著であれば，対象物を背景から並列に分離することができる。しかし，より混雑した複雑な視覚シーンでは，異なる物体を識別するために，再帰的で連続的な処理が必要になる(Lamme and Roelfsema, 2000)。連続的な処理とは，限られた注意力を画像内のある場所から別の場所に移動させることである。連続処理は，限られた注意資源を画像内のある場所から別の場所に移動させることであり，隠蔽的または顕在的な空間的注意の移動という形で行われる (Buschman and Miller, 2009)。図形と地面の分離や物体の識別には，視覚系の再帰的な接続，すなわち，同じ視覚野の近くのニューロンからの水平方向の接続と，より高い視覚野のニューロンからのフィードバック接続が役立つ。脳がどのようにして低レベルの特徴をグループ化し，一貫した物体識別を行うのかという問題は，1世紀近くにわたって研究されてきた。グループ化には，特に新規の物体や複雑な物体に対して注意が必要であると考えられている (Roelfsema and Houtkamp, 2011)。このことは，複数の特徴の組み合わせによって定義される物体を探す必要がある視覚探索課題において，特に重要であると考えられる。

神経学的には，対象ベースの注意効果は，対象の部分が精神的にトレースされることで，空間にゆっくりと広がっていく (Roelfsema et al, 1998)。対象物の外側に注意を移すことは，同じ距離にある対象物の内側に注意を移すことよりも大きなコストがかかるようである (Brown and Denney, 2007)。また，視覚対象に注意が向けられると，その対象のさまざまな特徴に対する特徴ベースの注意が視野内で活性化されると考えられている (O'Craven et al, 1999)。

もう 1 つの注意は，特徴的な次元全体に注意を向けることで，特徴的な注意と呼ばれることがある。例えば，ストループ検査では，色の名前が異なる色のインクで書かれており，被験者はその言葉を読むか，インクの色を言う必要がある。ここでは，あらかじめ特定の特徴に注意を向けることはできず，単語や色という次元にのみ注意が向けられる。神経学的には，次元の切り替えは視覚経路の感覚コーディングに影響を与え，前頭葉で制御されているようである (Liu et al, 2003)。

2.2.3. 視覚的注意の計算モデル

視覚的注意は，注意の神経科学において最も研究されているテーマの一つであり，注意がどのように機能するかについての多くの計算モデルを刺激してきた。一般に，これらのモデルは，様々な神経生理学的知見を統合して，注意の行動上の影響がどのように生じるかを説明するのに役立つ (Heinke and Humphreys, 2005)。

顕著性を計算するためのいくつかの計算モデルが考案されている (Itti and Koch, 2001)。これらのモデルは，低レベルの視覚特徴検出器 (通常，視覚系のものと同じように設計されている) を用いて，画像固有の 顕著性地図を作成し，同じ画像に対する人間のサッケードパターンを予測することができる。また，情報理論的な第一原理に基づいて顕著性を計算する別のアプローチも検討されており，ある種の視覚探索行動を説明することができた (Bruce and Tsotsos, 2009)。

注意の行動と神経の相関関係は，注意がボトムアップであってもトップダウンであっても同様のものがある。注意のバイアス競争モデルでは，刺激は神経反応を支配するために互いに競争する (Desimone, 1998)。注意 (ボトムアップまたはトップダウン) は，この競争を注意の対象である刺激に向けて偏らせることで機能する。偏り競争モデルは，単に直感を導くための「言葉モデル」として使われることもあるが，これを明示的に計算機上で具体化したものも作られている。トップダウンバイアスと，水平結合を介した局所的な競合を含む視覚経路の階層モデルは，注意の複数の神経効果を再現することができた (Deco and Rolls, 2004)。
同様の原理でスパイキングニューロンを用いたモデルも実装された (Deco and Rolls, 2005)。

同様のモデルは，上述のストループ検査のような属性命名課題を扱うために明示的に構築されている。例えば，選択的注意モデル (Selective Attention Model, SLAM) は，感覚符号化モジュールと運動出力モジュールの両方に局所的な競争があり，ストループ検査のような簡単な検査と難易度の高い検査における応答時間の既知の特性を模倣することができる (Phaf et al., 1990)。

視覚は，ベイズ推論の問題として組み立てられ，モデル化されている (Lee and Mumford, 2003)。この文脈では，注意は，推論がより困難な状況下で，主に事前知識を調整することによって不確実性を解決するのに役立つ (Rao, 2005)。例えば，Chikkerur ら (2010) では，空間的注意は物体の同一性に関する不確実性を低減するように機能し，特徴的注意は空間的不確実性を低減するように機能している。これらの原理は，注意の行動的特徴と神経的特徴の両方を捉えることができ，生物学的にインスパイアされた神経モデルに実装することができる。

注意の特徴類似性利得モデル (FSGM) は，特徴領域と空間領域の両方に適用できる，トップダウン注意の神経効果の説明である (Treue and Trujillo, 1999)。このモデルでは，注意によってニューロンの反応がどのように変化するかは，そのニューロンのチューニングに依存するとしている。チューニングとは，ニューロンが異なる刺激に対してどのように反応するかを示すもので，FSGM によれば，例えば青という色を好む (つまり強く反応する) ニューロンは，トップダウンで青に注目するとその活動が増強されることになる。また，FSGM では，好まない刺激に注意を向けると発火が減少し，増加しても減少しても，活動は注意によって乗算されるとされている。当初は計算モデルとして定義されていなかったが，その後，このような形の神経変調は，困難な視覚課題の成績を高めるのに有効であることがモデリングによって示されている (Lindsay and Miller, 2018)。

他のモデルでは，注意をネットワークを介した情報の動的なルーティングとして概念化している。この形式の注意の実装は，Selective Attention for Identification Model (SAIM) (Heinke and Humphreys, 2003) に見られる。ここでは，注意は網膜から「注意の焦点」とみなされる表現に情報をルーティングする。現在の課題に応じて，網膜表現の異なる部分が注意の焦点にマッピングされる。

2.2.4. 他の感覚モダリティにおける注意

聴覚における選択的注意の必要性を示す有名な例として「カクテルパーティー問題」がある。複数の話者や他の雑音で混雑した部屋の中で，一人の話者の音声に集中することが難しいという問題である (Bronkhorst, 2015)。この問題を解決するためには，ピッチなどの音声の低レベルの特徴を用いて，どの聴覚情報をさらに言語処理に回すかを決定する「早期」選択が必要だと考えられている。興味深いことに，選択的聴覚注意は，聴覚処理の最も初期のレベルである蝸牛でさえも神経活動を制御する能力を持っている (Fritz et al, 2007)。

体性感覚においても，空間的注意と特徴的注意が研究されている。体のさまざまな場所でタップを期待する手がかりを与えられた被験者は，その手がかりが有効であった場合，感覚をよりよく検出することができる。しかし，これらの効果は，視覚系での効果に比べて弱いようである (Johansen-Berg and Lloyd, 2000)。検出課題において，被験者が指上の刺激の向きを手がかりにした場合，反応時間が速くなる (Schweisfurth et al, 2014)。

キューイングされた味を検出する能力をテストした研究では，有効にキューイングされた味は，無効にキューイングされた味よりも低濃度で検出できることが示された (Marks and Wheeler, 1998)。これは，特徴に基づく視覚的注意で見られる行動効果を模倣したものである。嗅覚の特徴に対する注意については，あまり詳しく調べられていないが，視覚的に誘導された香りに対する期待は，その検出を助けることがある (Gottfried and Dolan, 2003; Keller, 2011)。

また，複数の感覚信号の統合を必要とする課題を実行するために，注意を複数のモダリティに分散させることもできる。一般的に，一致する複数の感覚信号を使用することは，単一のモダリティのみに頼る場合と比較して，物体の検出を助ける。興味深いことに，他の領域からの信号が同じように有効であっても，人間は視覚領域にバイアスをかけている可能性を示唆する研究もある (Spence, 2009)。具体的には，手がかりの空間的な位置を特定する必要がある課題では，視覚領域が最も優位になるようである (Bertelson and Aschersleben, 1998)。腹話術では，ダミーの口が動いているという視覚的な手がかりが，声源の真の位置に関する聴覚的な証拠よりも優先されることがよくわかる。また，視覚的証拠は，例えば，ゴム腕錯覚のように，触覚的証拠よりも優先されることがある (Botvinick and Cohen, 1998)。

感覚処理のクロスモーダルな性質のもう一つの効果は，あるモダリティでの注意の手がかりが，別のモダリティでの注意の方向付けを引き起こすことである (Spence and Driver, 2004)。一般に，手がかりのないモダリティでの注意効果は弱くなる。このようなクロスモーダルな相互作用は，内因性 (トップダウン) の注意と外因性 (ボトムアップ) の注意の両方の文脈で起こる可能性がある。

2.3. 注意と実行制御

複数の課題が同時に競合する場合，どの課題をいつ実行するかを決定する中央管理者が必要となる。さらに，課題をどのように実行するのがベストなのかは，履歴や文脈によって異なる。感覚入力と過去の知識を組み合わせて，効率的なタスクの選択と実行という仕事のために，複数のシステムを調整することが実行制御の役割であり，この制御は通常，前頭前野と関連している (Miller and Buschman, 2014)。前述したように，トップダウンの視覚的注意の源もまた，前頭葉領域に位置している。注意は，実行制御の出力として合理的に考えることができる。したがって，実行制御システムは，注意の対象を選択し，それを実行するシステムに伝えなければならない。前述の逆階層理論によると，上位領域が入力を得た領域に信号を送り，その信号が下位の領域に送られる，ということになる (Ahissar and Hochstein, 2000)。つまり，それぞれのポイントで，注意の指示を，対象となる領域にとって意味のある表現に変換する必要がある。このようなプロセスを経て，実行制御領域のハイレベルな目標は，例えば初期の感覚処理などの非常に具体的な変化をもたらすことができる。

また，過去の情報を活用したり，現在の目標を維持したりするためには，ワーキングメモリが必要であることから，実行制御とワーキングメモリは相互に関連しています。さらに，ワーキングメモリは，前頭前野の持続的な活動として認識されることが多い。実行制御，ワーキングメモリ，注意の三者間の関係の結果として，課題にとって望ましくない場合でも，ワーキングメモリの内容が注意に影響を与えることがある (Soto et al, 2008)。例えば，ある物体をワーキングメモリに保持しながら，同時に別の物体を視覚的に検索しなければならない場合，検索配列に記憶された物体が存在すると，検索に悪影響を及ぼす可能性がある (Soto et al, 2005)。このことは，ワーキングメモリが注意の実行制御を妨害する可能性を示唆している。しかし，注意の実行制御には，ワーキングメモリだけでは妨害されない要素がまだあるようである。このことは，被験者が記憶した項目を報告する必要があると思っているのに，代わりに出席した項目の検索配列を見せられると，視覚探索の成績がさらに低下するという研究に見られる (Olivers and Eimer, 2011)。このことは，ワーキングメモリ内のすべての対象が注意に何らかの影響を及ぼす可能性がある一方で，実行制御者はどの対象が最も影響を及ぼすかを選択できることを示唆している。

感覚モダリティの中で注意を柔軟にコントロールするだけでなく，モダリティ間で注意をシフトさせることもできる。行動実験によると，感覚モダリティ内の 2 つの異なる課題の間で注意を切り替える (例えば，視覚的な対象の位置を特定することから，それを識別することへ)，あるいは感覚モダリティ間で注意を切り替える (聴覚的課題から視覚的課題へ)と，計算コストがかかることがわかっている (Pashler, 2000)。このコストは通常，課題を切り替えた直後の試行と，同じ課題を繰り返している試行とで，成績がどの程度低下するかで測定される。興味深いことに，モダリティ内での課題の切り替えは，モダリティ間の切り替えよりも大きなコストがかかるようです(Murray et al., 2009)。同様の結果は、応答のモードを切り替えた場合にも見られ (例えば，ボタン押し 対 言語報告)，これは感覚処理に特有のものではないことが示唆されている (Arrington et al, 2003)。このような知見は，モダリティ内での切り替えには，同じ神経回路の再構成が必要であり，単に異なる感覚系の回路を働かせるよりも困難であることに起因すると考えられている。これは，異なる感覚システムの回路を単に動かすよりも難しいことである。効率的な実行制御者は，注意を移すことを決定する際に，このようなコストを認識し，理想的には最小化しようとする必要がある。スイッチコストは訓練によって低減できることが示されている (Gopher, 1996)。

注意の実行制御に関する最後の疑問は，学習によって注意がどのように変化するかということである。眼球運動の研究によると，検索された項目は，新規のものよりも馴染みのある設定でより迅速に検出できることが示されており，以前に学習した連想が顕在的な注意を導くことが示唆されている (Chun and Jiang, 1998)。このような利点は，海馬に依存していると考えられている (Aly and Turk-Browne, 2017)。しかし一般的には，注意を向ける方法の学習は，注意過程の他の側面ほど研究されていない。いくつかの研究では，被験者が無関係な課題情報を抑制する能力を高めることができ，その抑制の一般性はトレーニング手順に依存することが示されている (Kelley and Yantis, 2009)。注意学習の神経相関を見ると，イメージングの結果から，学習に伴う神経の変化は感覚経路そのものではなく，むしろ注意制御に関連する領域で起こることが示唆されている (Kelley and Yantis, 2010)。研究は必ずしも容易ではありませんが，乳幼児期や小児期における注意システムの発達は，注意がどのようにして学習されるのかについて，さらなる手がかりを与えてくれるかもしれません (Reynolds and Romano, 2016)。

2.4. 注意と記憶

注意と記憶には，さまざまな相互作用が考えられる。例えば，記憶の容量が限られているのであれば，脳が何を記憶に入れるかを選択することは理にかなっている。このようにして，全情報のサブセットを動的に選択する注意の能力は，記憶システムのニーズによくマッチしている。逆に言えば，特定の記憶を思い出すことは，限られた資源をどのように使うかという選択である。したがって，記憶の符号化と検索の両方が注意に頼ることができる。

記憶の符号化における注意の役割はかなり強いようである (Aly and Turk-Browne, 2017)。情報が適切に記憶に符号化されるためには，その情報が注意の対象であることが最適である。
被験者に単語のリストを記憶させながら，同時に注意を分散させる二次的な課題に従事させると，後でそれらの単語を意識的に思い出す能力が損なわれる (ただし，その単語を馴染みのあるものとして認識する能力はそれほど影響を受けない; Gardiner and Parkin, 1990)。二次課題の難易度を上げると，左腹下前頭回と前海馬における記憶符号化に関連する活動パターンが弱まり，背外側前頭前野と上頭頂領域における二次課題情報の表現が増加することが，イメージング研究で示されています (Uncapher and Rugg, 2005)。したがって，符号化という課題に置かれた限られた神経処理能力がなければ，記憶は損なわれます。注意もまた，空間的に定義された記憶の符号化に関与しており，場所細胞の表現を安定化させるようである (Muzzio et al, 2009)。

暗黙の統計的学習は，注意によってもバイアスがかかる。例えば，Turk-Browne ら (2005) では，被験者は赤と緑の図形からなる刺激の流れを見ていた。課題は，注意した色図形が 2 回連続して現れたときに検出することであった。被験者には知らされていなかったが，刺激の流れには統計的な規則性があり，3 つの図形が近接して出現する可能性があった。被験者は，3 つの形のセットを2つ見せられたとき，1 つは実際に共存する 3 つの形，もう 1 つは同じ色の形をランダムに選択したもので，実際の 3 つの形の方が親しみやすいと認識した。しかし，被験者は，参加した色の 3 つの図形である場合にのみ，本物の 3 つの図形をよりよく知っていると認識した。

しかし，意識的な注意を払わなくても，学習が行われることがある。例えば，Watanabe (2003) では，患者は視野の中心にある文字を検出する課題に取り組み，背景には閾値以下のコントラストでランダムドットの動きが表示されていた。この動きは，現在表示されている文字と相関のある方向に 10%のコヒーレンスを持っていた。この課題を学習する前後に，被験者は閾値以上の方向分類課題を行った。この課題を学習した後，方向分類は，対象となる文字に関連する方向に対してのみ向上した。これは，ターゲットによって活性化された報酬関連信号が，刺激の非注目要素についての学習につながったことを示唆している。

多くの行動研究では， 記憶の想起にどの程度の注意が必要かが検討されている。 例えば， 記憶した単語のリストを思い出すと同時に， カードを並べ替えるなどの二次的な作業をさせることで， 記憶の想起が， 作業と同じ限られた注意資源から得られるかどうかを調べることができる。 このような研究の中には， 注意を必要とする課題を同時に行うと記憶の回復が阻害されるというものがあり， 記憶の回復が注意に依存した過程であることを示唆している。 しかし， 正確な所見は，使用する記憶課題と非記憶課題の詳細に依存している (Lozito and Mulligan, 2006)。

もし， 記憶の検索が共有された注意資源から得られるものではないとしても， ある瞬間に他の記憶よりも鮮明に検索される記憶が選択されることは明らかである。 そのためには， 選択のプロセスが必要である。 神経イメージングの結果を見ると， 注意のトップダウン的な配分とボトムアップ的な捕捉を司る頭頂部の脳領域が， 記憶想起の際にも同様の役割を果たしていることが示唆されている (Wagner et al, 2005; Ciaramelli et al, 2008)。

記憶検索の研究は通常， 中長期記憶を対象としているが， ワーキングメモリ内の項目に注意を向けるメカニズムも提案されている (Manohar et al.2019)。 それは， ワーキングメモリの 2 つの異なるメカニズムに依存している。 非注意項目のためのシナプスの痕跡と， 注意項目のための持続的な活動である。

記憶の中には， 感覚処理の流れそのものの中で自動的に行われるものもある。 プライミングとは， ある時点での刺激の存在が， その後の刺激の処理や解釈に影響を与えるという心理学の有名な現象である。 例えば， 「school」という単語よりも 「hospital」という単語の方が 「doctor」という単語よりも早く認識される場合がある。 このように， プライミングには， 過去の刺激が現在の刺激に影響を与えるための暗黙的な記憶が必要である。 概念的プライミングや意味的プライミングに関するいくつかの研究では，プライミング効果が生じるためには， 最初の刺激に対する注意が必要であることが示されている (Ballesteros and Mayas, 2015)。 これは， 一般的に記憶のエンコーディングには注意が必要であるという知見と同じである。

ほとんどのプライミングはポジティブなもので， ある時点で刺激が存在すると， その刺激や関連する刺激の検出や処理が後から行われる可能性が高くなることを意味する。 このように， プライミングはボトムアップの注意を偏らせるものと考えることができる。 しかし， トップダウンの注意がネガティブプライミングを引き起こすこともある。 ネガティブプライミングでは， 前回の試行で妨害刺激として機能していた刺激が， 今回の試行で注意の対象となると， 成績が低下する (Frings et al, 2015)。 これは， 現在のターゲットとなる刺激に対して， 妨害刺激抑制のメカニズムがまだ活性化されているホールドオーバー効果に起因すると考えられる。

適応は暗黙の記憶の一形態とも考えられる。 ここでは， 同じ刺激に繰り返しさらされると， 神経の反応が低下する。 繰り返しの反応が減少することで， 刺激の変化がより顕著になる。 注意は， 注意を受けた刺激に対する神経応答を増加させることで， 適応の効果を打ち消す (Pestilli et al, 2007; Anton-Erxleben et al, 2013)。 このように， プライミングでも適応でも， トップダウンの注意は， ボトムアップの注意を導いている可能性のある下位レベルで起こる自動プロセスを克服することができます。

3. 機械学習における注意

人工的な注意という概念は， 現在の人工ニューラルネットワークの復活以前にも出てきているが， 現在よく使われているのは， ANN を中心としたものである (Mancas et al, 2016)。 人工ニューラルネットワークにおけるアテンションメカニズムの使用は， 脳における注意の必要性が明らかになったように， ニューラルシステムをより柔軟にする手段として生まれた。 機械学習における注意メカニズムは， 1 つの訓練された人工ニューラルネットワークが複数課題や， 長さ， 大きさ， または構造が変化する入力を持つ課題で優れた性能を発揮することを可能にする。 機械学習における注意の精神は，確かに心理学から着想を得ているが， その実装は， 後述するように， 生物学的な注意について知られていることとは必ずしも一致しない。

もともと ANN 用に開発された注意の形では， 注意機構は， エンコーダ-デコーダの枠組みで， 系列モデルの文脈で機能していた (Cho et al, 2015; Chaudhari et al, 2019)。 具体的には， 入力系列がエンコーダー (リカレントニューラルネットワークと思われる) を通過し， デコーダー(これもリカレントニューラルネットワークと思われる) の仕事は別の系列を出力することになる。 エンコーダとデコーダをつなぐのは，注意機構である。

一般的に， 符号化器の出力は，入力系列の各要素に対して 1 つずつのベクトルのセットです。 注意は， これらのベクトルのうち，どのベクトルを使って出力を生成すべきかを決定するのに役立つ。 出力系列は 1 要素ずつ動的に生成されるため， 注意は各時点で異なる符号化されたベクトルを動的にハイライトすることができる。 これにより， 符号化器は入力系列の最も関連性の高い部分を柔軟に利用することができる。

注意機構の具体的な仕事は， 符号化されたベクトル (v^i) のそれぞれに 1 つずつ， 一連のスカラー重み付け α_t^i を生成することである。 各ステップ t において，注意メカニズム (ϕ) は， デコーダの前回の隠れた状態 (h_{t-1}) と符号化されたベクトルに関する情報を取り込み， 正規化されていない重み付けを生成する。

$$\widehat{\alpha}_t = \phi\left(h_{t-1}, v\right) \tag{1}$$

注意は限られた資源であるため，これらの重み付けは相対的な重要性を表す必要がある。 α 値の合計が1になるように， 正規化されていない重み付けをソフトマックスにかける。

$$\widehat{\alpha}_t^i = \frac{\exp\left(\alpha_t^i\right)}{\sum_j \exp\left(\alpha_t^j\right)} \tag{2}$$

これらの注意値は， 符号化されたベクトルをスケーリングして， 符号化器が条件付けできる単一の文脈ベクトルを作成する。

$$c_t = \sum_j a_t^j v^j \tag{3}$$

このような注意の仕方は， 完全に微分可能なので， 単純な勾配降下法でネットワーク全体をエンド・ツー・エンドで訓練することができる。

このような AI 的な注意は， 反復的な再重み付けの一形態である。 具体的には， 前処理された入力の異なるコンポーネントを， 出力生成に必要に応じて動的にハイライトする。 これにより， 生物学的な注意と同様に， 柔軟で状況に依存したものとなっている。 そのため， 本質的に動的なものでもある。 系列モデリングにはすでに時間的な要素が含まれているが， この形式の注意は静的な入力と出力にも適用することができ (後に画像処理の文脈で説明します)， その結果， モデルにダイナミクスが導入される。

注意を払わない従来の符号化と復号化の枠組みでは，符号化器は，入力の長さや特徴に依存しない固定長のベクトルを生成し，復号化の過程では静的なものとなっていた。このため，長い配列や複雑な構造を持つ配列は，短い配列や単純な配列と同じ次元で表現されることになり，復号化は復号化過程で入力のさまざまな部分に質問することができなかった。しかし，入力を入力配列と同じ長さのベクトルの集合として符号化することで，復号化の各時点で関連する入力配列の部分に選択的に注目することが可能になる。ここでも，脳における注意の解釈と同様に，AI システムにおける注意は，限られた資源を柔軟に活用する方法として役立つ。復号化は入力のすべてを条件とすることはできないので，ある時点でボトルネックを導入しなければならない。注意のないシステムでは，固定長の符号化ベクトルがボトルネックになっていた。注意機構が追加されると，復号化が入力のどの部分に注目するかを決定する際に，ボトルネック (文脈ベクトルの形) が動的に生成されるため，復号化はより大きくなる。

このような注意機構を人工システムに追加する動機は，もちろんその性能を向上させるためである。しかし，注意のもう一つの利点は，解釈可能性であると言われている。復号化の過程で，入力のどの部分に注意が向けられているか (すなわち，どの α^i 値が高いか) を特定することで，復号化器がなぜそのような出力を出したかを理解できるかもしれない。ただし，注意の出力を解釈する際には，期待通りにモデルの振る舞いを説明できない場合もあるため，注意が必要である (Jain and Wallace, 2019; Wiegreffe and Pinter, 2019)。

以下の節で，この一般的な注意の概念の具体的な応用例と，この枠組みにうまく収まらない例を取り上げる。また，生物学との類似性にも注目する。

3.1. 自然言語処理における注意

以上のように，系列を処理するモデルには，注意機構が頻繁に追加されている。自然言語処理 (NLP) は，系列モデリングの最も一般的な応用分野の一つである。また，機械学習における注意のオリジナルの領域ではなく，生物学との共通点も少ないが，NLP は注意の最も一般的な応用分野の 1 つでもある (Galassi et al, 2019)。

人工ニューラルネットワークにおけるこのような形の注意の初期の応用例として，翻訳課題があった (Bahdanau et al, 2014; 図3)。この仕事では，リカレントニューラルネットワークが入力文を，文中の各単語に 1 つずつ対応する「注釈」ベクトルのセットとして符号化する。出力であるターゲット言語の文は，リカレントニューラルネットワークによって一度に 1 つの単語ずつ生成される。生成された各単語の確率は，それまでに生成された単語，リカレントニューラルネットワークの隠れ状態，および注意機構によって生成された文脈ベクトルの関数である。ここで，注意メカニズムとは，出力ネットワークの隠れ状態と，現在の注釈ベクトルを取り込み，すべての注釈ベクトルに対する重み付けを行う小さなフィードフォワードニューラルネットワークである。

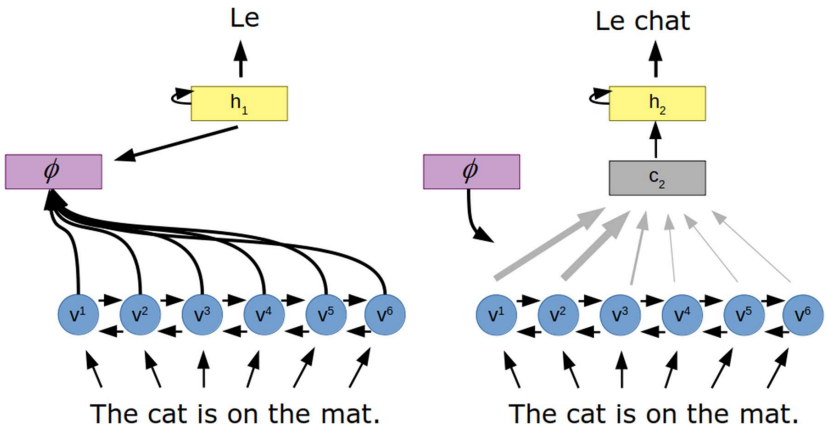

図 3 ニューラル機械翻訳への注意
翻訳されるべき文は，リカレントニューラルネットワークを介して一連のベクトル (v) に符号化される。注意機構 (ϕ) は，デコーダの隠れた状態 (h) とこれらのベクトルを用いて，エンコードされたベクトルをどのように組み合わせて文脈ベクトル (c) を生成すべきかを決定する。このベクトルは，デコーダの次の隠れた状態に影響を与え，したがって翻訳文の次の単語に影響を与える。

このようにして文中のすべての単語の情報をブレンドすることで，ネットワークは出力する単語を生成する際に，前の部分と後の部分のどちらかを利用することができる。これは，標準的な語順が異なる言語間の翻訳に特に有効である。著者らは，入力文の中で注意が適用された場所を視覚化することで，注意がこの問題に役立つことを確認した。

この最初のアプリケーション以来，言語翻訳のための注意ネットワークの多くのバリエーションが開発されてきた。Firat ら (2016) では，1 つの言語だけでなく，複数の言語のペア間の翻訳に使用できるように，注意機構が適応された。Luong ら (2015) では，著者らは，すべての入力単語に一度にアクセスする能力が必要かどうかを判断するために，異なる注意の構造を探っている。また，Cheng ら (2016) では，より柔軟に文の表現を作るために，文の符号化と復号化を行うリカレントニューラルネットワークに注意機構を追加した。

2017年，影響力のある「Attention is All You Need」という論文では，機械翻訳にまったく異なるスタイルのアーキテクチャが活用された (Vaswani et al, 2017)。このモデルは再帰性を持たないので，学習がよりシンプルになる。代わりに，文中の単語が並列に符号化され，これらの符号化によってキー表現とクエリ表現が生成され，これらが組み合わされて注意の重み付けが作られる。これらの重み付けは，単語の符号化自体をスケーリングして，モデルの次の層，つまり「自己注意」として知られる過程を作る。
この過程は繰り返し行われ，最終的には自己回帰復号化と相互作用する。自己回帰復号化は，コード化された入力 (標準的な注意の形) と以前に生成された出力に柔軟に焦点を合わせることができる注意機構も備えている。この新しい注意機構に与えられた名前である Transformer は，これまでの多くのモデルを凌駕し，すぐに機械翻訳だけでなく他のタスクの標準となった (Devlin et al, 2018)。

興味深いことに，自己注意は，もともと機械翻訳に使われていたリカレント注意モデルに比べて，生物学的注意との共通点が少ない。
第一に，脳は言語処理や注意選択などの逐次処理課題において必然的に再帰性に依存しているのに対し，自己注意は再帰性とダイナミクスの役割を減らしている。第二に，自己注意は単語間の水平方向の相互作用を提供する。これにより，符号化された文の中の単語は周囲の文脈の中で処理されるが，この機構には，復号化器のニーズによって駆動される明らかなトップダウンの要素は含まれていない。実際，自己注意は，ある状況下では，画

像処理で頻繁に使用される標準的なフィードフォワード計算であるコンボリューションを単純に実装することが示されている (Andreoli, 2019; Cordonnier et al, 2019)。 このように， 自己注意は， 限られたリソースに基づいて課題固有の注意のような選択を行うよりも， 良い符号化を作成することが重要である。 時間的課題の文脈では， 心理学における最も近い類似物はプライミングかもしれない。 プライミングは， 先行刺激に基づいて後続刺激の符号化を変化させるからである。 もちろん， 機械学習エンジニアの直接的な目標は， 脳を再現することではなく， 課題をうまくこなせるように簡単に訓練できるネットワークを作ることである。 このような制約があるため， 機械学習が大きく進歩しても， 必ずしも脳に近いモデルができるとは限らない。

人間の言語処理における注意の研究は， 神経科学の他の分野ほど大規模ではないが， 読書中の眼球運動を追跡する研究がいくつか行われている (Myachykov and Posner, 2005)。
彼らは， 現在読んでいる文章を明確にするために， 特に代名詞の先行詞を見つけるという文脈で， 文章の前の部分を振り返ることを発見した。
このような注意のシフトは， 現在の処理要求に最も関連する過去の情報を示している。

3.2. 視覚課題のための注意

神経科学や心理学と同様に， 機械学習の研究の大部分は視覚的な課題で行われている。 コンピュータビジョンの注意喚起ツールの 1 つである「顕著性マップ」は， エッジ， 色， 奥行きなどの低レベルの視覚的特徴に基づいて， 画像内のどの領域が最も顕著であるか， またその領域が周囲とどのように異なるかを識別するものである (Itti and Koch, 2001)。 このようにして， 人や動物の「ボトムアップ」の注意によってどの領域が捉えられるかを示すのが， 顕著性 saliency マップである。 コンピュータ科学者は， 画像処理パイプラインの一部として， さらなる処理のために領域を識別するために， 顕著性地図を使用している。

より近年では， コンピュータビジョンモデルは深層学習に支配されている。 そして， 2012 年の ImageNet チャレンジで成功して以来 (Russakovsky et al, 2015)， 畳み込みニューラルネットワークは， 機械学習における視覚課題のデフォルトアーキテクチャとなっている。

畳み込みニューラルネットワークのアーキテクチャは， ほ乳類の視覚システムに大まかに基づいている (Lindsay, 2020)。 各層では， フィルターのバンクが下層の活動に適用される (第 1 層では 画像)。 これにより， 神経活動の H×W×C のテンソルが作成される。 チャンネル数 C は適用されたフィルターの数に等しく， H と W はフィルターを適用した結果の 2 次元特徴地図の高さと幅を表す。

畳み込みニューラルネットワークにおける注意は， 分類， セグメンテーション， 画像を利用した自然言語処理など， さまざまな課題成績を向上させるために用いられている。 また， これらの注意過程は， 神経科学の文献と同様に， 空間的な注意と特徴に基づく注意に分けられる。

3.2.1. 空間的注意

NLP 課題の注意に使われる構造を基に， 視覚的注意が画像キャプションに適用されている。 Xu et al. (2015)では， 符号化モデルは畳み込みニューラルネットワークである。 注意機構は， 第 4 の 畳み込み層での活動に作用する。 キャプションの各単語が生成されると， 画像表現の空間的な位置に渡る重み付けの異なるパターンが作られる。 このようにして， キャプション生成時の注意は， 翻訳課題における符号化された単語ベクトルのセットを， 符号化された画像位置のセットに置き換える。 このモデルは， 高い重みを持つ場所を視覚化することで， キャプションを生成している現在の単語に最も関連する対象に注意するように見える。

このスタイルの注意は， 空間的な位置に対する視覚的な特徴の重み付けされた組み合わせを生成するため， 「ソフト」と呼ばれている (図4B)。 「ハード」な注意は， 他のすべてを犠牲にして復号化に渡す 1 つの空間位置を選択する代替形態である (図4A)。 Xuら（2015）では， どの場所がこのハード注意を受けるべきかを決めるために， 各空間場所に対して生成された注意の重みを確率として扱った。 これらの確率に応じて 1 つの場所が選択される。 この確率的な要素をネットワークに加えることで， 訓練はより難しくなるが，ソフトな注意よりもやや良好なパフォーマンスが得られることがわかった。

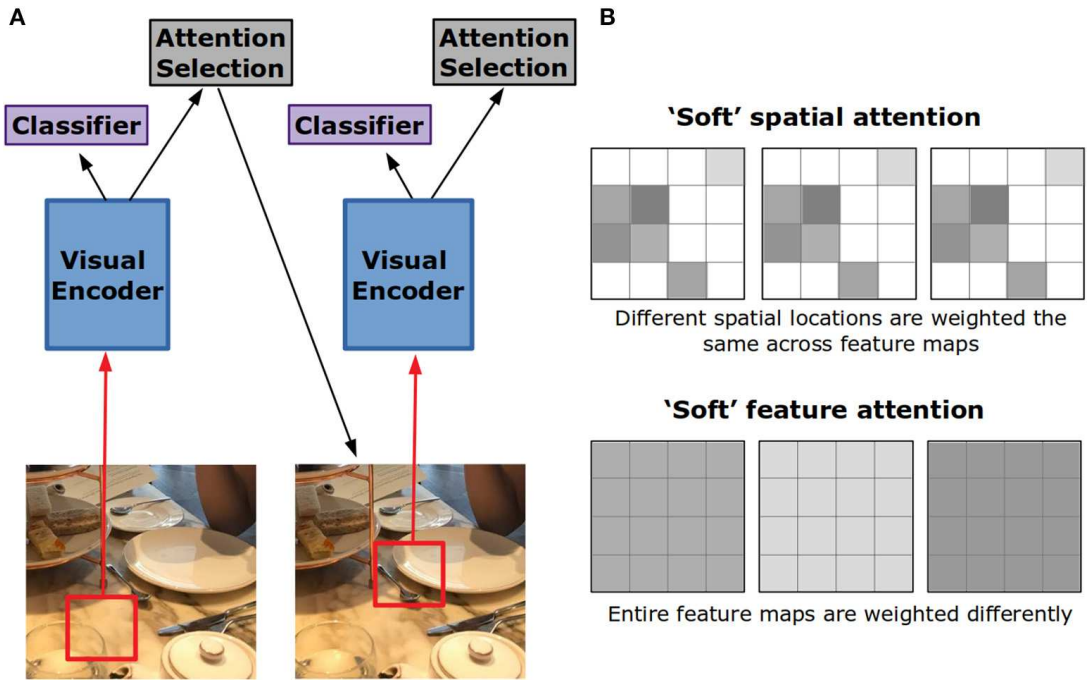

図 4 人工ニューラルネットワークにおけるハードな視覚的注意とソフトな視覚的注意。 (A) ハードアテンションでは， ネットワークは画像全体のごく一部からしか入力を得られない。 この部分は， 注意選択機構を介して， ネットワークによって反復的に選択され

る。 入力が中心窩にあれば， ネットワークは低解像度の周辺部を使ってこの選択を導くことができる。 (B) 畳み込みニューラルネットワークの特徴地図は， 下層にフィルタを適用することで作成される活性化の 2 次元グリッドである。 ソフトな空間注意では， このグリッド上の異なる位置が異なる重みを持つ。 ソフトな特徴注意では， 異なる特徴地図が異なるように重み付けされる。

2014 年の研究では， 強化学習を用いてハードな注意ネットワークを訓練し， 厳しい条件下で物体認識を行うことができた (Mnih et al, 2014)。 このモデルの中核となるのはリカレントニューラルネットワークで， ネットワークが行う複数の「チラ見 glimpses」で取り込んだ情報を記録するとともに， 次のチラ見の位置を出力する。 チラ見するたび， ネットワークは画像の小さなパッチから中心窩のような入力 (中心部は高解像度， 周辺部は低解像度で表現される) を受け取る。 ネットワークは， これらのチラ見から得られた情報を統合して， 画像内の物体を見つけ出し， 分類しなければならない。 これは上述のハードな注意と似ているが， ここでの場所の選択は， 画像のどの部分を次にサンプリングするかを決定する (上のケースでは， すでに処理された画像のどの場所を復号化に渡すかを決定していた)。 これにより， ネットワークは画像のすべてを処理する必要がなくなり， 計算機資源を節約することができる。 また， 画像内に複数の物体が存在し， ネットワークがそれぞれを分類しなければならない場合にも有効である (Ba et al 2014)。 最近の研究では， 事前訓練のステップを追加することで， 複雑な画像に適用されるハードな注意の成績が向上することが示されている (Elsayed et al, 2019)。

生物学的な注意と人工的な注意の対応は， 多くの点で， 視覚的な空間的注意の場合に最も強く現れる。 例えば， 画像のさまざまな場所を順次サンプリングして処理するハードな注意は， サッケードの過程を再現したもので， 神経科学や心理学の文献にある「オバートな視覚的注意」に似ている。 ソフトな注意は， ネットワークへの入力に変化を与えることなく， ネットワークの画像表現の異なる領域を動的に再重み付けするという点では， 隠れた空間的注意に似ている。 また， ソフトな注意の適用方法は， 特定の場所にあるすべてのユニットの活動を乗算的にスケーリングすることであるため， 隠蔽された空間的注意に関する神経学的な知見を再現している。

ソフトな空間的注意は， 視覚的な質問と回答 (Chen et al, 2015; Xu and Saenko, 2016; Yang et al, 2016) や， 動画でのアクション認識 (Sharma et al, 2015) など， 他の課題にも使用されている。 また， ハードな注意は， 実体セグメンテーション (Ren and Zemel, 2017) や，異なるレベルの画像解像度を用いて適用した場合のきめ細かな分類 (Fu et al, 2017) にも用いられている。

3.2.2. 特徴への注意

ソフトな空間的注意の場合，重みは画像表現の異なる空間的位置で異なるが， その位置にあるすべての特徴チャンネルでは同じである。 つまり， 異なる視覚的特徴を表すネットワーク内のユニットの活動は， 画像空間内の同じ場所を表していれば， すべて同じように変更されることになる。 特徴的注意は， 個々の特徴地図を動的に再重み付けすることで， 特徴の処理を空間的に大域的に変化させることを可能にする。

Stollenga et al. (2014) では，畳み込みニューラルネットワークに特徴ベースの注目メカニズムを搭載している。 画像が標準的なフィードフォワードアーキテクチャを通過した後， ネットワークの活動は， 異なる層の異なる特徴地図をどのように重み付けすべきかを決定するポリシーに渡される。 この再重み付けは， 異なるネットワーク活動をもたらし， それが異なる再重み付けにつながる。 ネットワークがいくつかのタイムステップで実行された後，最終層の活性が画像内の対象を分類するために使用される。 重み付けの値を決定するポリシーは強化学習によって学習され， 事前に学習された任意の畳み込みニューラルネットワークに追加することができる。

Chen et al. (2017) のモデルは， 特徴的な注意と空間的な注意を組み合わせて， 画像のキャプション付けを支援する。 畳み込みネットワークのフィードフォワードパスの活性は， 以前に生成された単語とともに注意機構に渡され， CNN の各層で異なるチャネルに対する注意の重み付けを作成する。 これらの重みは活性値をスケーリングするために使用され， 次に別の注意機構が空間的重みを生成するために同じ手順を行う。 空間と特徴の両方の注意重みが生成され， 各時点でネットワークに適用される。

De Vriesら（2017）のモデルでは， 質問の内容を利用して，視覚的な質問と回答の課題のために CNN が画像を処理する方法を制御する。 具体的には，言語埋め込みネットワークのアクティビティを多層パーセプトロンに通して， CNN の各チャネルのバッチ正規化のための加算パラメータを生成する。 この手順は条件付き一括正規化と呼ばれ， 質問に依存した特徴の注意の形として機能する。

動的な特徴の再重み付けの異なる形が「squeeze-and-excitation」ネットワークに現れる (Hu et al, 2018)。 このアーキテクチャでは， 異なるチャネルに適用される重み付けは， 同じ層の他のチャネルの活動の非線形関数である。 前述の「自己注意」と同様に，これは， 重み付けがネットワークの後半の活動の関数であり， かつ/または 出力生成器のニーズによって偏っている， より「トップダウン」なアプローチとは根本的に異なる。 生物学的には， このような相互作用は， 分割正規化などの計算を行うことが知られている視覚野内の水平結合に最もよく似ている (Carandini and Heeger, 2012)。

特徴に基づく注意の生物学的研究では， 被験者は通常， 特定の視覚的特徴に注意を向けたり検索したりするように合図される。 この方法では， 注意すべき特徴は事前に知られており， 特定の副次課題に関連している (例えば， 一般的な形状検出課題の所定の試行で特定の形状を検出する)。 これは， 上記の人工的な特徴注意の例とは異なり， 特定の画像に関する知識が得られる前に， 外部の手掛かりがネットワーク処理を偏らせることはない。 むしろ， 特徴の再重み付けは画像自体の機能であり， 一定の課題におけるネットワークの成績を向上させることを目的としている (これは， 前述の人工的な空間的注意の形態の場合にも当てはまる)。

生物学的注意の研究にキューパラダイムを用いる理由は， 実験者が注意をどこに置くかをコントロールできる (つまり， 知ることができる) からである。 しかし， 明示的な手がかりがなくても， 私たちの脳は常にどこに注意を置くかを決定していることは明らかである。 これらは， 視覚系への局所的および長距離的なフィードバック接続によって媒介されていると考えられる (Wyatte et al, 2014)。 したがって， 生物学的特徴注意の研究と AI システムでの使用とでは課題構造が異なるが， この違いは表面的なものに過ぎないかもしれない。 基本的に， AI システムは， トップダウンの情報を手がかりの形で与えられるのではなく， フィードフォワードの画像情報を使って， トップダウンの注意信号を内部で生成している。

とはいえ， 人工的なシステムの中には， 外部からキューされた特徴に注意を払うことができるものもある。 例えば， Cao et al. (2015) のネットワークでは， カテゴリに対する事前処理を設定することで， 特定のカテゴリのローカライズがうまくいくようになっている。 Wang ら (2014) のネットワークは， 畳み込みではないが， 特定の物体カテゴリの検出を偏らせる手段を持っている。 また Lindsay and Miller (2018) では， 手がかり付き物体検出課題中の生物学的特徴の注意のいくつかの性能と神経的側面をCNN を用いて再現した。 Luo ら (2020) では，CNN で手がかり付き注意の形式を使うことのコストとメリットが検討された。

上述したように， 活動の乗法的スケーリングを使用することは， 生物学的な視覚的注意から得られたある知見と一致している。
さらに， 特徴地図全体を同じスカラー値で変調することは， 特徴注意が視覚系において空間的に大域的に作用するという前述の知見と一致する。

3.3. マルチタスク注意

マルチタスク学習は，機械学習の中でも特に難しいテーマである。例えば，物体分類，エッジ検出，顕著な領域の特定を行う CNN のように，1 つのネットワークに複数の異なる課題を要求された場合，個々の課題を実行するために必要な重みが互いに矛盾してしまうため，学習が困難になる。ひとつの方法は，課題固有のパラメータセットを用意して，共有ネットワークの活動を課題ごとに異なるように調整することである。常にそう呼ばれるわけではないが，ネットワークの機能を柔軟に変化させることから，これは合理的には注意の一形態と考えられる。

Maninis ら (2019) では，共有されたフィードフォワードネットワークを複数課題のすべてで学習させる一方で，課題に特化したスキップ接続やスクイーズ・励起ブロックは，その特定の課題でのみこの活動を変調するように学習させている。これにより，ネットワークは，全課題に共通する処理を共有することで利益を得つつ，それぞれの課題に多少特化することができる。

Rebuffiら (2017) では，同様の手順で，複数の異なる画像ドメインで分類を行うネットワークを作成した。ここでは，入力画像からドメインを特定し，課題固有のパラメータセットを実行時に自動的に選択することができた。

Zhaoら（2018）では，同じ画像をネットワークに渡して，異なる次元 (例えば，写真の人物が笑っているかどうか，若いか古いか) に沿って分類することができる。これらの異なる分類を実行するために，特徴チャネルの課題固有の再重み付けが使用される。

Strezoskiら（2019）のモデルでは，ハードな特徴注意の一形態と解釈できるものを用いて，課題ごとに異なる情報をルーティングしている。特徴チャンネル上の 2 値化マスクは，課題ごとにランダムに選ばれる。これらのマスクは，全課題での訓練中とランタイムに課題固有の方法で適用される。このネットワークでは，課題固有の注意パラメータは学習されない。なぜなら，これらのマスクは事前に決定され，訓練中に固定されるからである。代わりに，ネットワークは，異なる課題を実行するために，結果として生じる異なる情報経路を使用することを学習する。

最近の研究では，課題固有のパラメータの概念は完全に取り払われた (Levi and Ullman, 2020)。その代わりに，フィードフォワード CNN の活性化を課題入力と組み合わせて 第 2 の CNN に通し，調整用ウェイトのフルセットを生成する。これらの重みは，元のネットワークの活動をユニット固有の方法で調整する (これにより，空間と特徴の両方に注意を払うことができる)。その結果，複数の視覚課題に柔軟に対応できる単一のフィードフォワード結合荷重セットが得られる。

同じ入力を異なる課題に応じて異なるように処理する場合，これらのネットワークは本質的に，特徴的な注意に依存したモード内課題スイッチングの一形態を実施していることになる。この点では，先に述べたストループ検査に最も似ているかもしれない。

3.4. 記憶への注意

深層ニューラルネットワークは，明示的な記憶を持たない傾向にあり，そのため記憶への注意は研究されていない。しかし，Neural Turing Machine は，外部メモリストアを含むハイブリッドなニューラルアーキテクチャである (Graves et al, 2014)。ネットワークは，訓練を通じて，これらの記憶保持と効果的に相互作用する方法を学習し，記憶された系列のソートや反復などの課題を実行する。この相互作用を容易にすることが，注意の一形態である。記憶は，ベクトルの集合として保存されている。この記憶から情報を取り出すために，ネットワークは各ベクトルに重みを生成し，記憶の加重和を計算する。この重みを決定するために，リカレントニューラルネットワーク (外部から課題に関連した入力を受け取る) は，ベクトルを出力し，記憶はこのベクトルとの類似性に応じて重み付けされる。このようにして，各時点で，ネットワークは文脈に関連した記憶にアクセスすることができる。

前述したように，脳がどのようにして注目すべき記憶を選択し，それに注目するのかは完全には解明されていない。このモデルでは，類似性計量尺度を使用しているため，神経科学の文献にある連想記憶モデルと同様に，生成された活性ベクトルとの重なりに基づいて記憶が検索されることになる。これは，後者の問題，つまり記憶への注意が脳内でどのように実行されるのかという問題に対する機構を提供するものである。このモデルでは，生成された活性ベクトルが，どのような記憶に注意を向けるかを制御しており，生物学との関係はあまり明確ではない。

4. 人工的な注意と生物学的な注意の間の将来の相互作用のアイデア

これまで述べてきたように，生物学からのヒントを得て，AI ニューラルネットワークではすでにいくつかの注意の例が生まれている (図5 にまとめた)。このような注意メカニズムを追加することで，これらのシステムの性能はかなり向上したが，まだまだ不十分な点も多く，さらなるインスピレーションを得る機会も存在している。近い将来，このようなインスピレーションは，現在存在する特殊な人工システムを少しずつ改良していく形で得られるだろう。しかし，脳にインスパイアされた AI の真の目的は，より統合された，多くの課題に柔軟に対応できる多目的エージェントを実現することである。

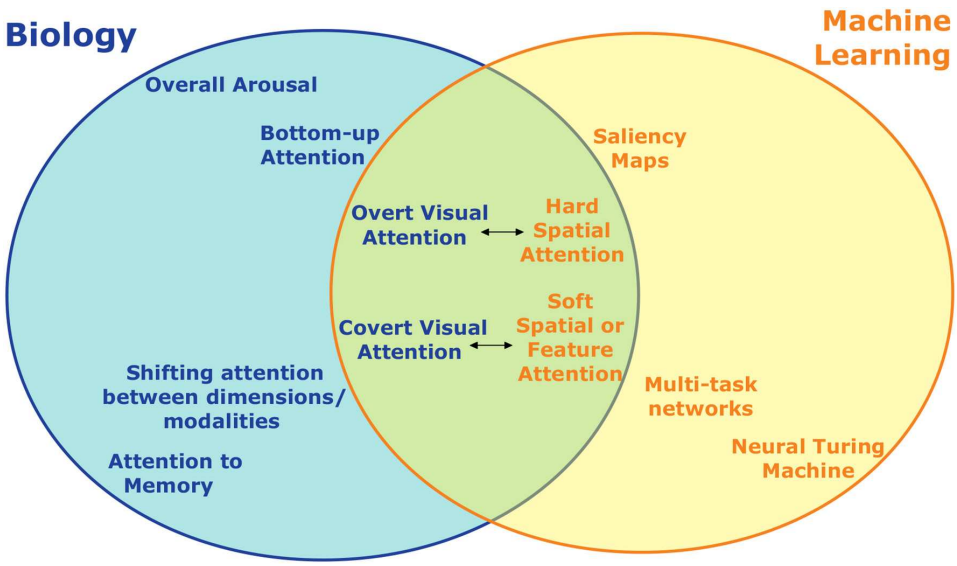

図 5 神経科学・心理学と機械学習で研究されているさまざまな注意の種類とその関連性を不完全ながらまとめたもの。
左は生物学的に研究されている注意の区分， 右は人工知能や機械学習のために開発された注意の区分である。 同じ水平方向の位置にあるトピックは， ある程度の類似性があり， その距離が類似性の近さを示す。 例えば， 視覚的注意の形態は， 生物学と機械学習の間で最も重複しており， 最も直接的に比較することができる。 全体的な覚醒度など， 注意の形態によっては， 明らかに人工的な類似性がないものもある。

4.1. パフォーマンスを向上させるには

脳における注意の働きの研究には， 表裏一体ともいえる 2 つの要素がある。
1 つ目は， 注意がどのようにして成績を向上させるのか， つまり， 注意に関連する神経の変化がどのようにして脳を課題遂行能力に優れたものにするのかという問題である。 2 つ目は， 注意がどのようにして， また，なぜそのように展開されるのか， つまり， ある特定の項目や課題が注意の対象として選択され， 他のものは選択されないのはどのような要因によるのか， という問題である。

神経科学者たちは， 前者の疑問について多くの時間を費やして研究してきた。 しかし， これらの知見を AI 神経システムに適用するのは， 大部分が簡単ではないかもしれない。 活動の乗法的スケーリングは， 生物学的にも人工的にも現れるものであり， 注意を実装するための有効な手段である。 しかし， 脳内で観察される注意の効果の多くは， 主に， ノイズの多いスパイク状のニューロンが伝える信号を増加させる手段として意味をなしている。 これには， ニューロン間の同期の増加や， 発火のばらつきの減少などが含まれる。 このような変化を示すアナログがディープニューラルネットワークになければ， そこからヒントを得ることは難しい。 さらに， ニューラルネットワークの学習手順は， 明確に定義された課題成績を向上させるために必要な活動の変化を自動的に決定することができるため， 生物学的変化から得られる教訓はあまり意味がないかもしれない。

一方で， 注意が活動電位の高さやバースト性， 正確なスパイク時間などのスパイク特有の特徴に影響を与えるという観察結果は， スパイキングネットワークの有用性を示していると考えられる。 具体的には， スパイキングモデルでは， 注意が制御できる自由度が高いため， 注意がより大きな， あるいはより微妙な影響を与える可能性がある。

注意の解剖学的構造を観察することは， AI システムのアーキテクチャを設計する人々に有益な洞察を与えるかもしれない。 例えば， 視覚的注意は， V4 などの後期視覚領域でより強く活動を調節するように見える (Noudoost et al, 2010)。 だが， 聴覚的な注意は， 処理経路のかなり早い段階で活動を調節することができる。 このように， 注意がどのレベルで作用するかは， アーキテクチャの変数として関連している可能性がある。 このような観点から， 最近の研究では， Transformer モデルの初期層から自己注意を取り除くことで， 特定の自然言語処理課題での成績が向上し， また， 言語処理中の人間の fMRI 信号の予測にも適したモデルになることが示されている (Toneva and Wehbe, 2019)。

ある感覚モダリティで注意を引くと， 別のモダリティでも同じ物体や場所に注意が展開されるというクロスモーダルなキューの存在は， 異なる感覚システム間の直接的な相互作用をある程度示していると考えられる。 機械学習におけるマルチモーダルモデルの多くは， 完全に別々の処理経路を使用し， 最終的にのみ結合されるが， 異なる入力経路間の水平方向の接続を可能にすることで， それらの処理を調整することができる。

また， 注意は， 通常の感覚処理で行われるような適応と相互作用する。 一般に， ニューラルネットワークモデルには適応のメカニズムがない。 つまり， 同じ入力を複数の時間ステップで与えられた場合， ニューロンは活動を低下させる手段を持たない。 適応は， 変化や異常を目立たせるのに有効であることを考えると， 適応させることは有用であると考えられる。 適応のあるモデルでは， 注意機構は， 繰り返される刺激が重要であると判断された場合， 適応したニューロンを再活性化するように働くはずである。

最後に， 注意の形態の中には， 同じシステムに対して複数の働きをするものがあるようである。 例えば， 視覚的注意は次のような働きをすると考えられている: (1) 大脳皮質の視覚ニューロンの活動を調節することで， その感度を高める。 (2) 大脳皮質下の活動を変化させて， 感覚情報が異なる形で読み出されるようにする(Birman and Gardner, 2019; Sreenivasan and Sridharan, 2019)。 このようにして， 注意は， 脳の異なる部分にある 2 つの異なるメカニズムを使って， その効果を生み出す。 注意がモデルアーキテクチャの複数の構成要素を補完的に調節できるようにすることで， より強固で効果的な影響を与えることができるかもしれない。

4.2. 注意をどのように配置するか

複雑で統合された人工知能を作り出すためには，注意をどのように展開するかという問題の方が重要であると考えられる。流れてくる刺激の中から関連する情報を選択したり，取り組むべき最適な課題を選んだり，あるいは何かに取り組むかどうかを決定するには，エージェントが自分の状態，環境，ニーズを統合的に理解している必要がある。

生物学的注意から影響を受ける最も直接的な方法は，それを直接模倣することである。例えば，スキャンパスモデルは，何年も前から顕著性の研究に存在している。これらは，人間が画像を見ているときに行う一連の固視を予測しようとするものである (Borji and Itti, 2019)。注意を訓練するためのより直接的なアプローチは，Linsley ら (2018) で用いられた。ここでは，被験者に物体分類に最も関連する画像の領域をラベル付けさせることで，人間のトップダウン注意の大規模なデータセットを収集した。この方法で作成された課題固有の顕著性地図は，物体認識を主課題とする深層畳み込みニューラルネットワークの注意を訓練するために使用された。その結果，この方法で中間層の活動に影響を与えると，成績が向上することがわかった。また，Zagoruyko and Komodakis (2016) では，教師の顕著性地図を学習する別の方法が示されている。

課題と人間の視覚領域から収集した神経データを組み合わせた訓練も，CNN の性能向上に役立っている (Fong et al, 2018)。特に注意課題で収集した神経データを使えば，注意モデルの訓練に役立つ可能性がある。このような転送は，他の課題も行うことができる。例えば，読書中の眼球運動を追跡することで，NLP モデルに情報を与えることができる。これまでのところ，眼球運動は品詞タグ付けモデルの訓練に使われている (Barrett et al, 2016)。興味深いことに，乳児は周りの大人が何に注意を向けているかに気を配ることから学ぶことができ，より広くエージェント間で注意を調整することは，社会的な種において非常に有用であると考えられる。したがって，他者の注意は，注意がどのように導かれるかに影響を与えるはずである。共同注意を調整する試みは，注意システムに統合する必要があるだろう (Kaplan and Hafner, 2006; Klein et al, 2009)。

興味深いことに，乳児は周りの大人が何に注意を向けているかに注意を向けることから学ぶことができ，より広範にエージェント間で注意を調整することは，社会的種族において非常に有用であると考えられる。したがって，他者の注意は，注意の導き方に影響を与えるはずである。共同注意を調整する試みは、注意システムに組み込まれる必要がある（Kaplan and Hafner, 2006; Klein et al, 2009)。活動は，いくつかの可能な目標のうち，いつでもどの目標を達成すべきか，したがってどこに注意を置くべきかを柔軟に決定する必要があるだろう。この問題は，強化学習，特に副次課題の選択を伴う階層型強化学習の問題と密接に関連している。このような決定は，予想される正または負の結果に基づいて行う必要がある。実際，注意と報酬の間には密接な関係があり，以前に報酬を得た刺激は，もはや報酬を得られない状況でも注意を引きつける(Camara et al, 2013)。人間がどのかだいをいつ実行するかをどのように選択するかについて理解を深めれば，人間の行動をマルチタスク AI の設計に反映させることができる。

そのためには，異なる過程を制御する脳の限られた能力の配分は，その制御の期待値に基づいて行われるという，Shenhav ら (2013) が提唱した理論が役に立つかもしれない。この枠組みでは，背側前帯状皮質は，制御の認知的コストを含む多様な情報を統合して，制御の期待値を計算し，その結果，注意などの過程を指示する役割を担っている。複雑な課題における人間の実行制御を理解するための別のアプローチとして，逆強化学習がある。この方法は最近，人間が使用する報酬関数とポリシーを決定するために，視覚探索中の眼球運動のデータセットに適用された (Zelinsky et al, 2020)。

生物学的な注意を促進するが，人工的な注意システムではおそらく十分に表現されていない追加的な要因は，好奇心である (Gottlieb et al, 2013)。生物学的には，新規の刺激，混乱した刺激，驚くべき刺激が注意を引くことがあり，下側頭葉と末梢皮質は，馴染みのある入力に対する反応を減少させる適応メカニズムを介して，新規の視覚的状況を通知すると考えられている。状態の値の推定の一部に新規性を含める強化学習アルゴリズムは，このような探索を促すことができます (Jaegle et al, 2019)。しかし，異なる状況で驚きや新しさを具体的にどのように計算するかは，必ずしも明確ではない。生物学的注意に関するこれまでの研究では，ベイズ的に驚きや情報収集の観点から注意選択を理解しており，こうしたフレーミングは AI システムにも有用であると考えられる (Itti and Baldi, 2006; Mirza et al, 2019)。

注意の選択における最後の問題は，葛藤をどのように解決するかである。覚醒，ボトムアップ，トップダウンなど，脳には複数の注意の形態があるが，適切な注意の所在に関する葛藤はどのように解決されるのだろうか。視覚系を見てみると，これらの複数のシステムが対象とする局所回路がこの課題を担っているように見える。これらの回路は，神経調節のための入力とトップダウンの信号を受け取り，それらを活動を促すボトムアップの入力と統合しなければならない。水平方向の接続がこの競争を仲介し，勝者と勝者の関係を築くメカニズムを利用することができる。これは，AI システムのアーキテクチャでも真似ることができる。

4.3. 注意と学習

注意は，何が記憶に入るかを決定する役割を果たし，学習を導く。注意を備えた AI システムの多くは，訓練中に注意機構を含んでいる。このようにして，注意機構は基本アーキテクチャと一緒に学習される。しかし，Neural Turing Machine を除いては，注意機構が機能した後，モデルは学習を続けない。したがって，これらのシステムでは，学習や記憶を制御する注意の能力はまだ明示的に考慮されていない。

注意は，入力中の関連する成分や関係性に学習を向けることで，データを効率的に利用するのに役立つだろう。例えば，顕著性地図は，様々なコンピュータビジョン課題の前処理の一部として使用されている (Lee et al, 2004; Wolf et al, 2007; Bai and Wang, 2014)。本質的に顕著な領域のみに後続の処理を集中させることで，無関係な領域への無駄な処理を防ぐことができ，また，ネットワーク学習の観点からは，これらの領域へのオーバーフィッティングを防ぐことができる。しかし，このように顕著性マップを利用するには，問題に応じた顕著性の定義が必要である。人間のボトムアップの注意を引く画像の特徴を利用することは，コンピュータビジョンの問題ではうまくいっているが，他のモダリティの人間のデータを見ることも有用である。

これに関連して，乳幼児の研究では，顔などの関連する刺激に注意を向けるための事前知識を持っていることが示唆されている。このような事前知識を用いることで，重要な刺激をどのように処理するか，また，関連する特徴にどのように注意を向ければよいか，という両方の学習を起動させることができる (Johnson, 2001)。

トップダウンの注意は，データのどの部分を処理するかを決めることに加えて，処理中にネットワークのどの要素を最も働かせるべきかを選択することと考えることができる。ネットワークのどの部分が最も強く働いているかで学習が行われるので，これも注意が学習を導く手段の一つである。任意の入力に応じて更新されるパラメータの数を制限することは，ドロップアウトやバッチ正規化の使用に見られるように、正則化の効果的な形態である。注意は，どのユニットを使うかをランダムに選択するのではなく，この課題の成績にも役立つユニットを選択するように制約されている。そのため，より課題に特化した形での正則化となる。

このように，注意は，ネットワークが既に学習した他の課題成績を乱さないようにしながら，特定の課題でより良い成績を発揮するようにネットワークをアップデートすることを目的とする継続的学習に特に役立つ可能性がある。関連する概念である条件付き計算は，最近，継続的学習の問題に適用されている (Lin et al, 2019)。条件付き計算では，ネットワークのパラメータは現在の入力の関数であり (そのため，注意によって行われるタイプの変調の極端な形と考えることができる)，ネットワークを効率的な継続学習のために最適化するには，ネットワークの効率的な継続学習のためにネットワ

ークを最適化するには， 異なる入力間の干渉の量を制御する必要がある。 より一般的には， 注意は， 望ましくないシナプスの変化を防ぐための手段であると考えるのがよいだろう。

注意と学習もループしている。 具体的には， 注意は世界について何が学習されるかを導き， 内部世界モデルは注意を導くために使用される。 この相互依存関係は， 近年， 人間の学習や意思決定の説明に成功している認知ベイジアン推論モデルも取り入れた強化学習フレームワークの観点から形式化されている (Radulescu et al.2019)。 大脳基底核と前頭前野の相互接続は， 強化学習と注意選択の相互作用を支えていると考えられている。

より抽象的なレベルでは， 脳のアーキテクチャに注意が存在するだけで， 表現学習に影響を与えることができる。 意識の大域的ワークスペース理論では， どの瞬間にも， 脳の活動から選択された限られた量の情報がワーキングメモリに入り， さらなる共同処理に利用できるとしている (Baars, 2005)。 これに触発されて， 機械学習における「意識先行」は， 基礎となる高次元の状態表現に適用される注意から生じる低次元表現を持つ神経ネットワークアーキテクチャを強調している (Bengio, 2017)。 この低次元表現は， 将来の状態を要約したり予測したりするのに使えるような， 抽象的なレベルで世界を効率的に表現する必要がある。 この注意を媒介としたボトルネックの存在は， 行動を導き， 予測を行うために柔軟に組み合わせることができるように， すべてのレベルで分離された表現を促すというトリクルダウン効果をもたらす。

楽器の演奏など， 多くの複雑なスキルの学習には， 意識的な注意が必要である。 しかし， いったん学習が完了すると， これらの処理は自動的に行われるようになり， 注意力が解放されて他のことに集中できるようになる可能性がある (Treisman et al, 1992)。 この変換のメカニズムは完全には解明されていないが、タスクの負担を別の， おそらくより低い/より反射的な脳領域に移動させることに依存しているようである限り， 注意によって異なる働きをすることができる複数の冗長な経路を持つことは， 人工システムにとって有益であると考えられる (Poldrack et al, 2005)。

4.4. 注意の限界 バグか機能か？

生物学的注意は完全には機能しない。 前述したように， 異なる種類の注意を切り替えるときにパフォーマンスが低下したり， ピークパフォーマンスに達するためには覚醒レベルがちょうど良くなければならなかったり， トップダウンの注意が無関係だが顕著な刺激によって中断されたりする。 AI システムに注意を移す際の問題は， これらの制限を避けるべきバグなのか， それとも取り入れるべき機能なのかということです。

気が散りやすいというのは， 一般的には， 注意力のバグというよりも， むしろ特徴のように思われる。 課題に集中しようとしても， 生命を脅かす可能性のある環境の変化を意識し， 気が散ってしまうことは有益なことである。 問題となるのは， 脅威でもなく， 関連性のある情報でもない入力に対して， エージェントが過度に注意をそらす場合です。 したがって， 人工的なシステムでは， トップダウンの注意力の強さのバランスをとり， 予期しないが情報を提供する刺激の処理を可能にする必要がある。 例えば， 注意の瞬きとは， ターゲットとディストラクタの流れの中で， 第 1 のターゲットの後に 第 2 のターゲットがすぐに現れた場合に， 被験者がそのターゲットを見落としてしまう現象のことである (Shapiro et al, 1997)。 これは成績を低下させるが， 脳が最初のターゲットを処理して行動する時間を与えるためには必要なことかもしれない。 このようにして、 フォロースルーを確実にするために気が散ることを防ぐ。

人工的なものであれ， 生物学的なものであれ， どのようなエージェントであっても， そのエネルギー資源にはある程度の限界がある。 そのため， いつ社会に出るのか， それとも睡眠などの省エネ状態に入るのかを慎重に判断することが常に重要になる。 多くの動物では， 睡眠はスケジュールに沿って行われるが， 前述したように， 注意を要する状況では睡眠が遅れたり中断されたりすることもある。 そのため， いつ睡眠状態に入るかは， 起きていることで何が得られるかという費用対効果の分析に基づいて決定しなければならない。 睡眠には記憶の定着など， 省エネ以外にも重要な働きがあることが知られており， その判断は複雑なものになるかもしれない。 AI システムがこの判断を下すためには， 現在の状態と将来の要求を統合的に理解する必要がある。

5. 結論

注意は， 心理学， 神経科学， 人工知能にまたがる大きくて複雑なテーマである。 この名前で研究されているテーマの多くは， その機構に重複はないが， 限られた資源を柔軟にコントロールするという中核的なテーマを共有している。 柔軟性や資源の賢明な利用に関する一般的な知見は， 人工知能の開発に役立つし， 特定の感覚モダリティや課題に注意を向けるための最適な方法に関する具体的な知見も得られる。