Title
The Neural Mechanisms of Visual Attention

Author
Noah, Sean

Publication Date
2021

The Neural Mechanisms of Visual Attention

By

SEAN LI MING NOAH
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

George R. Magnun, Chair

Joy J. Geng

Steven J. Luck

Committee in Charge

2021

Table of Contents

Abstract

Attentional selection mechanisms in visual cortex involve changes in oscillatory activity in the EEG alpha band (8 to 12 Hz) – decreased alpha indicating focal cortical enhancement and increased alpha indicating suppression. This has been observed for spatial selective attention and attention to stimulus features such as color and motion. Theoretical work suggests that the control of oscillatory neural activity may be a crucial component of the mechanisms of attentional signal propagation through cortex, and in particular, the modulation of alpha band activity might constitute an information routing mechanism across all of visual cortex. This hypothesis leads to the prediction that alpha band modulation should also accompany attention to high-level visual features, such as object categories. To test this prediction, we investigated whether attention to objects involves alpha-mediated changes in focal cortical excitability. We conducted three experiments to look for systematic differences in the pattern of EEG alpha power across the scalp, and assess whether any such systematic differences were driven by the engagement of object-based attention to different object categories. In Experiment 1, twenty volunteers (8 males; 12 females) were cued (80% predictive) on a trial-by-trial basis to different objects (faces, scenes, or tools). Support vector machine decoding of alpha power patterns revealed that late (>500 msec latency) in the cue-to-target foreperiod, only EEG alpha differed with the to-be-attended object category. In Experiment 2, to eliminate the possibility that decoding of physical features of the cues led to our results, twenty-five participants (9 males; 16 females) performed a similar task where cues were non-predictive of the object category. Alpha decoding was now only significant in the early (<200 msec) foreperiod. In Experiment 3, to eliminate the possibility that task set differences between the different object categories led to our Experiment 1 results, twenty participants (9 males; 11 females) performed a predictive cuing task where the discrimination task for different objects was identical across object categories. The results replicated Experiment 1. Together, these findings support the hypothesis that the neural mechanisms of visual selective attention involve focal cortical changes in alpha power for not only simple spatial and feature attention, but also high-level object attention in humans. These findings support the Specificity of Control model of attention, according to which a top-down source of attentional control signals issues instructions over individualized channels to sensory sites according to the particular receptive properties of those sites, and an identical mechanism to elaborate information processing operates at the sensory sites. We followed up the planned analyses with a series of ERP decoding analyses to corroborate our theoretical interpretation, and to provide evidence that alpha band modulation is directly related to an immediate perceptual benefit afforded by attention. Altogether, our findings suggest that the modulation of oscillatory neural activity in the alpha frequency band is an essential component of the enhancement mechanism of attention, operating throughout visual cortex.

Acknowledgements

Writing this dissertation was the most challenging and the most fulfilling thing I’ve ever done. But unlike free climbing El Capitan or swimming the English Channel – the kind of feats that are undertaken entirely to test the limits of human ability and thereby award a profound sense of achievement (neither of which endeavor I have ever attempted, obviously; I’m using my imagination here) – finishing my dissertation was fulfilling not because of how challenging it was. Instead, it seems to me the causal arrow was a little more convoluted, and ultimately pointed the other direction.

Knowing that this piece of writing would encapsulate everything I've come to learn and postulate about the cognitive neuroscience of attention bestowed me with a sense of purpose right from the get-go and confronted me with a high bar of expectation. I expected that this written document would fulfill a desire, deeply ingrained in me, to glimpse fundamental truths about human experience and the natural world using tools of science. Sure, my writing and the years of thought and research it documented would reveal an infinitesimally narrow slice of the human condition – but this slice would not be any less meaningful to me by its narrowness; the years of effort it represented would have been highly focused and all the more intense because of it. The expectation that I would accomplish all this, that I would satisfy such a deep and profound yearning to contribute something however tiny to humanity’s scientific enterprise, and in doing so add a novel element to our understanding of the human mind, set the bar for this work high. And that high bar made doing the work mortally challenging. Thus, however teleologically loopy it may sound, writing this dissertation was the most challenging thing I had ever done entirely because it was the most personally fulfilling thing I would have ever done.

With that said, let me now turn to the real reason for this section's existence: acknowledgment of those people who brought me to where I am now, guiding me toward this most meaningful culmination of my whole scholastic life. The sense of fulfillment that I cultivated in crafting this long scientific text may be personal, but the research described in the following pages was never a solitary undertaking. This science was a group effort from the very start, and I will be forever grateful for the mentors, the collaborators, the peers, the friends, the role models, and the multifarious other supporters who guided me through the odyssey of graduate study. As I sit down and write this section of the dissertation, as I reflect on everyone who helped to deliver me to where I am today and the myriad forms taken on by that help, I am nearly overwhelmed with gratitude.

I want to give everyone their due acknowledgement. There's no brevity required here, no exit music to usher me back to my seat – this is a piece of writing that can be perused or skimmed at one's leisure. But if I took the paper space to reflect the totality of support I've received, this book would be too unwieldy to handle, and then I would never get the final signatures I need to officially earn my degree. So, at the risk of being long-winded, I will acknowledge all the major figures in my life who contributed to this work, but with an eye to practical concerns I will make certain abbreviations, and thus my acknowledgements will be shrunken from the true magnitude of the feelings they convey.

I first want to express my gratitude for my faculty advisors. My major advisor is Dr. Ron Mangun, and his guidance over the years of my graduate schooling was absolutely instrumental for developing me into a functional cognitive scientist from a mere aspirant whose curiosity was as powerful as his knowledge was limited. Fiery passion without guardrails is a dangerous thing, and I can recall numerous periods over the last four years when I didn't know how to properly channel my driving curiosity, my wild ambition, or my intellectual recklessness and stubborn insistence to plunge headfirst into sidetracking topics and tangential lines of thinking. Ron's guidance was always wise and expertly administered, gentle when a light touch was all I needed to see things clearly, but never capitulating when I would act out. Ron has been, and will continue to be, my primary source of scientific inspiration and professional role model. I feel extremely fortunate to have been able to apprentice under his mentorship. I may never reach the heights of his career, but I know that whenever I'm faced with a tough decision or an intractable question, Ron's voice in my head will help me navigate whatever difficulties lie ahead.

Along with Ron, the faculty advisors that shaped the course of my graduate study most profoundly were Dr. Joy Geng and Dr. Steve Luck. Both of these professors taught me innumerable lessons (that I carefully internalized), challenged my assumptions, strengthened my scientific worldview, and ultimately made me grow as a researcher. Together with Ron, Dr. Geng and Dr. Luck formed my Individual Advising Committee, most of my Qualifying Exam Committee, and my Dissertation Committee. And yet these formal labels understate the role that these professors played in my life. Without the guidance of Ron, Joy, and Steve at every step of my graduate education, I would have been totally lost. My research interests would have remained aimless, never to have been refined into proper questions. My curiosity about the huge topics in our field would have been forever stunted, never to develop past the earliest stages of rigorous inquiry and understanding.

Two other professors who played major roles in my graduate education and development as a cognitive neuroscientist were Dr. Mingzhou Ding and Dr. John Henderson. Dr. Ding and the members of his lab at the University of Florida were crucial collaborators in the overarching scientific project undertaken by the Mangun lab during my time there, both in terms of the theoretical development and the methodological implementation of the research. Without the technical expertise provided by the Ding Lab, many of the research methods deployed in the course of this work would never have been possible. And while the Mangun Lab trainees were learning methods, the Ding Lab was plowing ahead on data collection and analysis. Our weekly joint virtual lab meetings thus would always give me plenty of raw data to chew on even while I was still struggling to understand the finer points of data collection and experiment design, and were crucial for the proper development of my analytic skillset and my ability to think and talk on my feet. More than anything else, my participation in weekly lab meetings has given me the practice and the confidence boost required for speaking about research extemporaneously in front of groups and clearly, effectively communicating scientific ideas. Moreover, Dr. Ding was a member of my Qualifying Exam committee, overseeing my ascendance from PhD student to PhD candidate. Dr. Henderson was also a member of that committee, and his input during my exam helped steer me toward new areas of theory and research results that I wasn't previously aware of.

All the faculty mentioned heretofore, and all the faculty of the UC Davis Psychology Department as a whole, deserve special additional acknowledgement for their handling of graduate education during the Covid-19 pandemic (not to mention undergraduate education, which must have been an entirely separate challenge). As was the case for many people, the year 2020 was undoubtedly the most difficult year of my life. As I survey the year from my current vantage point in 2021, I am overcome with gratitude for the accommodations made by the Psychology Department for the benefit of all its students. For those struggling with research stagnation, social isolation, financial hardship, family stress, career uncertainty, political bewilderment, and other related sources of existential anxiety, the empathy and compassion expressed by Department leadership was nothing short of therapeutic. At the time of this writing, the future is still uncertain, and the long-term effects of this pandemic era, personal, professional, and societal look grim. But whatever happens next, I won't forget how steady leadership and compassionate understanding kept things on track and bolstered a sense of normalcy that helped me cope with the challenges encumbering me. Ultimately I count myself extremely lucky to have been able to weather the storm of 2020 in the shelter of graduate school, where a delicate stability was maintained by the assiduous efforts of administrators, leadership, and peers, when such a nurturing and protective environment was hardly available to the rest of the world.

In that vein, I also want to express my earnest thanks to Angela Scully, Taleen Ledesma, Marie Burns, Theresa Geer, and all the other program administrators and staff members of the Psychology Department, the Center for Mind and Brain, and the Center for Neuroscience who I came in contact with in the course of my time at UC Davis. I also wish to sincerely thank all those staff members who I never had the chance to meet. Without their diligent administration, their timely reminders, and the innumerable other duties that they carried out behind the scenes, I'm certain that all my ambition to become a scientist would have fizzled on the damp fuse of my attention to the details of record keeping and the vital business of bureaucracy.

Next, I want to acknowledge everyone who supported me in ways other than direct mentorship. This kind of support was primarily emotional but not any less crucial to my advancement through the graduate program.

My friends deserve a special acknowledgement here at the outset. I bonded tightly with my Psychology cohort and I am now lucky enough to count some of my closest friends among my fellow graduate students. Meaningful friendship in my cohort gave me a "life" to counteract the "work," in the mold of the exalted "work-life balance." But also, throughout my grad school years, my friends were continual sources of intellectual inspiration. When I needed something to stir up my stagnant thought patterns, they would challenge me. And when I needed to commiserate, they could assuage my anxiety. Amazingly, they would put up with my most bewildering behaviors, my petulant and swinging moods, my impulses, and my most aggravating affectations. I would be nothing without my friends, but thankfully they know that already.

In the grand scheme of things, my friends have had it easy, though – my family has been dealing with me for 30 years. They're tired of me, without a doubt. But they have never wavered in their loving support for me and any of the foolhardy pursuits I choose to lunge face-first into. For that I will be eternally grateful to them, and now if only they could figure out how to cash my statements of undying gratitude, they could recoup some of the expenses they've incurred on my behalf. In the meantime, I leave them with the knowledge that I am indebted to them to a supernatural degree and I love them immensely. As I write this dissertation, I am burrowed into a closet in the basement of the house belonging to my sister and her husband, in the beautiful and neighborly Green Lake area in Seattle, Washington. Kelly and Max are the sort of generous souls who will let me stay with them – nominally it's a "writing retreat" – while I grind away at this capstone of my graduate education. And I'm the kind of selfish goblin who will utterly take advantage of their generosity and intrude into their lives for an entire month. My deepest thanks to them for giving me a space conducive to doing this focused work.

An entire dissertation in itself could be written to acknowledge the ways that I am grateful for my partner Sara. I will leave the bulk of that work to future scholars and make a long story short in the span of this paragraph. I met Sara as an undergraduate at UC Berkeley, in a student housing co-op where we both lived, having been drawn there by the life-organizing principles of vegetarianism. The year was 2012, and the world was supposed to be ending soon. Having then nothing to lose, Sara agreed to join me on a long road trip. We set out North on Highway 101 in the blue morning light on a clear day in December, and we’re still on that trip to this day. My life with Sara is an endless adventure and I couldn’t have possibly found a better accompaniment on this latest chapter of the journey.

Lastly, a sincere Thank You to all the members of the Mangun Lab past and present. From my first day in the lab on that sweltering summer day in 2017, to the fateful eighteenth of March, 2020, when Yolo County’s first shelter-in-place order went into effect, ending my residence at the Center for Mind and Brain, my experience as a graduate student researcher was greatly impacted by the interactions I had with the fellow members of my lab – post docs, grad students, and undergraduate research assistants alike. Being around people driven by the same wonderment at the mystery of human experience, the same specific interest in attention, and the same sense of purpose in life was both encouraging and humbling. If any of you are reading this, please consider the following 200 pages to be the cover letter of my application for employment in your lab.

Sean Li Ming Noah

January 2021

Foreword

The weather turned clear so I decided to take Phoebe for a walk. Nearly bulldozing herself through the front door, Phoebe was exuberant, as she often is when I awaken her from one of her long naps with an invitation to tow me into the forest. On this particular day she had already been sleeping for the entire morning and the first few hours of the afternoon, which for a dog isn't that unusual, but for this reason her energy reserves were full to bursting. And so we took the longer route, up into the wooded hills.

That afternoon the very atmosphere was practically narcotic. Billows of Autumn clouds were breaking and rolling upward at the seams, revealing expanding fissures of limpid blue sky. The air was warm, and wafting on a gentle breeze were the redolent butterscotch and vanilla perfumes of Jeffrey pines and Ponderosas. Insects stridulated softly in panorama, chipmunks squeaked as they bounced alongside us, and chickadees punctuated the soundscape with their miniature chirps. The alpine bucolia was in full display.

Phoebe and I ascended the steepest part of the hill and when the paved trail ended abruptly we continued onto a compacted dirt path and traipsed into the wilderness. We plodded carefully over big rocks here and there, and did our best to sidestep the prickly Ceanothus shrubs that became larger and more unavoidable as we pushed deeper into the forest. Here I say "we" as if Phoebe was an equally contributing partner in the effort not to snag the leash on a protruding branch, sprain an ankle, or suffer multiple puncture wounds inflicted by a mass of bristling thorns, but that's too generous. Phoebe was pulling me mercilessly with the full strength of her athletic 60-

pound frame, darting chaotically from side to side, with no heed for any hazard. She was ecstatic, as she generally is on these long walks in nature, but made even more so by dint of her overflowing energy stores. And despite the unremitting force from her leash and the challenge of keeping pace with her, I found myself slipping into a relaxed and genial mood. The weather was just that good. That, and being immersed in the natural splendor of a forest cradled amidst rolling slopes and jagged outcrops of granite, set me at perfect ease.

Then, in the thick of the forest, I heard the unmistakable sounds of a pack of coyotes. They weren't close: somewhere at least 100 feet away, somewhere deep in the woods and beyond the endpoints of all my lines of sight that radiated into the dense vegetal growth. The coyotes were yipping and yelling in their disconsonant manner that never fails to unnerve me completely. I paused and stilled my breath, bracing Phoebe's leash with two clenched fists. Up here in the mountains, coyotes are something that need to be worried about. Lone coyotes will occasionally snag little dogs, and I've heard stories from neighbors and friends about packs of coyotes taking down huskies, labs, things that size. For this reason we diligently avoid walking in the hours around dawn and dusk, when the coyotes are thought to be most active. I was surprised, and scared, to find myself in their presence on this afternoon walk.

And yet, that fear was tinged with excitement, for a reason that hopefully will clarify why I have chosen to open my dissertation with this anecdote. My fear was tinged with excitement because the sudden nearness of the coyotes caused my attention to come online, fully engaged, in a way that I, an aspiring attention researcher, couldn’t help but observe gleefully, like a naturalist catching sight of a rare animal stepping into a clearing.

As a student of the cognitive neuroscience of attention I have spent nearly the last decade absorbing as much knowledge as I could on the subject. From attention’s early and basic folk psychological definitions to the multifaceted modern construct formed in parallel across the disparate yet interconnected fields of cognitive psychology, neuroscience, neurology, philosophy, even computer science, I had delved deeply into any and all scientific characterization of attention that I could find, in order to better understand this vital but mysterious aspect of human mental experience. Over the course of my graduate study I participated in innumerable discussions about attention’s properties, its functions, its reason for existing, and whether it even existed at all. And yet, for all these academic examinations of the concept of attention, I had scarcely ever really understood it intimately, and I didn’t appreciate how clearly it could manifest. Encountering the coyotes in the woods that day, startling me to a full alertness and attentiveness projected in stark contrast against the background of my relaxed montane stroll, showed me vividly and undeniably all the properties of attention that I had learned about academically yet only experienced faintly in my everyday life.

Because of my years of narrow focus on the topic, I was keenly attuned to attention’s many qualities, and now watched them all unfold in my mind. I recognized them, like species out of a field guide, in their natural habitat: orienting both endogenous and exogenous, distractor suppression, guided visual search, template formation and matching, and feature enhancement.

The coyote pack, by their sounds, could be generally localized, and my eyes were locked toward that direction. With steady intent, and no lapse or discontinuity, I scanned the forest scene laid out before me. Someone could have snuck up behind me and pickpocketed my underwear right through my pants, and I wouldn't have noticed. As I scanned my field of view, seeking out any visual sign that the coyotes might be padding towards us, both the motive force guiding my eyes and my perceptual impression of the scene were tuned away from shapes and colors that couldn't plausibly signal coyotes. Tangled red brambles, black-scorched columns of bark, deep green leaves, and brittle stems splotched with verdigris lichen all disappeared from my awareness, while any form that might tenably be a stalking coyote was clarified and highlighted in the field of my awareness – every coarse dun patch, every curved line evoking a hunched ridgeline of feral canine shoulders, every miniscule movement that could have been a trotting leg or a swishing tail intercepted my eyes in their darting flight and was subjected to rapid and discriminating analysis. And in the realm of sound: My hearing sensitized to the noises of the coyotes, reflexively casting aside auditory input that could be categorically distinguished from any sound a coyote might make – all bird calls, all gurgling creeks lapsed out of my mind. A brush of needles against a mossy log carried with it a possibility of panting breath. A snapping twig was deafening and all-consuming. In that moment I was a machine with the singular purpose of finding coyotes, a sieve to isolate their presence from out of the tumultuous and ambiguous world of sensation.

Slowly Phoebe and I backed down the hilly forest path we had carved and safely got away from the coyote pack. We never faced any real danger, but the episode stayed with me. Experiencing attention's full force, imposed on me by something urgent, deep and instinctual, and simultaneously being able to catalog its features, renewed my appreciation for the power and central importance of attention in the mind's operation. Moreover, it reaffirmed my opinion that attention can be used like a wedge to breach some of the most seemingly impossible scientific questions concerning the relationship between the objective physicality of the brain and the subjective phenomenology of the mind, an opinion that I hope to expound over the course of my dissertation.

The specific topic that my dissertation research addresses is object-based attention. Object-based attention is a subtype of attention writ large, one among several taxonomically distinct variants that include spatial attention, feature-based attention, temporal attention, etc., but I believe that this object-based subtype is actually categorically distinct from other forms of visual attention in the context of laboratory research only. I believe that in the laboratory, methods of operationalization set limits on psychological phenomena – granted, these limits are necessary to put empirical pins in fundamentally subjective entities and occurrences – and our scientific terminology emerges from those operational limits. I'm interested in object-based attention because I think it hints at a much broader range of mental phenomena and neurophysiological processes than its laboratory treatment suggests, and at various points in course of this dissertation, I will explain my reasoning for this opinion in fuller detail.

The research recorded in this dissertation is a close examination of a specific hypothesis concerning the neural mechanisms underlying selective visual attention. I carried out this work over the course of about three experiments and several different analysis methods. The chapter immediately after this Foreword is an Introduction, providing general background on the scientific history of attention in psychology and cognitive neuroscience, a similarly high-level discussion on the topic of visual objects, and a more speculative and theoretical account of object-based attention that attempts to merge our understandings of attention and object perception into a coherent framework. The three chapters following the Introduction document the original research that I conducted over the course of my graduate study. The next chapter after those ones is a discussion of my original research results and a placement of those results into the larger theoretical context that I lay out in the Introduction. The short and final chapter is the Conclusion, where I summarize the research and theory-crafting I performed and reiterate the ideas that I most hope will be taken away from this body of work.

I hope that the narrow question posed by my experiments will be answered over the course of the relevant chapters to a satisfying degree. I also intend, in the wide-ranging discussions that precede and follow the chapters devoted more singularly to my experiments, to make the case for how the results of my research can be interpreted within a broader theory of attention. I hope that this argument will be compelling.

Chapter 1: Introduction

What is Attention?

In 1890, in his treatise *The Principles of Psychology*, William James opened a chapter on attention by writing that “every one knows what attention is.” Although this notoriously blithe assertion is sometimes challenged or rebuked (Brigard, 2012; Hommel et al., 2019; Buzsáki, 2020), it is striking how much of James’s chronicle of foundational attention research and exposition of attention’s qualities can be related to the scientific investigation of attention today.

Whether or not one agrees with James’s claim that everyone knows what attention is, his description of attention couches the phenomenon in familiar language. He defines attention as:

> *“…the taking possession of the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others.”*

James laid out several distinct qualities of attention: its immediacy to experience, its limited capacity, its exclusivity, and its ability to facilitate effective goal-oriented behavior. Over the course of the chapter, James identified that attention enhances or enables our sensory perceptions, internal conceptions, ability to discriminate between perceptions or ideations, entry into memory, and reaction times. The spirit of his statement that attention is familiar to everyone directly follows from how he conceptualized attention: Everyone knows what attention is because attention is how people know anything at all. Attention is the mental process by which ideas or perceptions are brought to conscious awareness. By merely experiencing a sensory impression of external objects or events, or forming a coherent train of thought, one is engaging attention. Thus, everyone knows what attention is because if one even pauses to consider whether one knows what attention is, one is using attention.

James was writing at a time when modern Psychology was in its early stages. His writings on attention were largely introspective and speculative, albeit with some references to empirical work such as the foundational efforts of Wilhelm Wundt on reaction time, codifying early notions of psychology into a formal science. Nevertheless, the collection of attention's qualities laid out by James closely resembles more formal and refined operational definitions of attention that are used in research settings today.

In contemporary lay usage, attention has multiple meanings. It is generally thought of as the ability to stay task oriented, as in cases when somebody is told to pay attention to their work, or when somebody is reprimanded for not dedicating sufficient attention to an important responsibility. This usage overlaps with the various psychiatric constructs of attention deficit disorder (ADD), according to which an inability to stay task oriented and the susceptibility to distraction by task-irrelevant stimuli is sufficiently harmful and interruptive to everyday life so as to be diagnosable as a mental illness. The public's widespread familiarity with ADD understandably reinforces the general understanding of attention as something primarily relegated to continuous task performance and task orientation.

But task orientation is not the primary quality of attention that cognitive neuroscience seeks to understand, although it does fall under the umbrella of the different ways that attention is studied in a formal scientific setting. Alongside the ability to be task oriented, researchers today define attention in a number of ways. One modern way of conceptualizing attention is as a psychological function that restricts cognitive operations to a subset of the information that could potentially be engaged at a given time, to elaborate or enhance the processing of that information with the purpose of improving the efficacy, accuracy, or speed of behavioral responses (Luck and Vecera, 2002). A corollary of this definition is that attention is needed because the capacity of mental processes is limited and unable to accommodate all available information at one time. Attention, as just defined, is the function that selectively engages with a subset of available and mutually-competing information to control and facilitate effective behavior (Desimone and Duncan, 1995).

The quality of improving processing of selected information has been subdivided into two components: the control processes that guide enhancement mechanisms toward the selected information, and the enhancement mechanisms themselves (Luck and Gold, 2008). This distinction is analogous to the operation of a spotlight to illuminate a targeted physical object: the process of aiming the spotlight is necessary for the illumination of the target, but conceptually and mechanistically distinct from the way that the target is made more visible by the light falling on it. These distinct processes are sometimes referred to as orienting and detecting, respectively (Posner, 1980).

The definition just given describes attention at the level of cognitive psychology. Attention researchers in cognitive neuroscience aim to link attention's mental qualities to underlying neural processes. Examining attention at the neural level typically begins with broad definitions such as one given by Buschman and Kastner: "Attention is the selective prioritization of the neural representations that are most relevant to one's current behavioral goals" (Buschman and Kastner, 2016). This definition introduces an essential quality of attention as being able to selectively operate over neural representations. This perspective emphasizes the selective nature of attention, and posits a mapping between the selective mental quality of attention, what James described as "taking possession of the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought," and a selective process over discrete neural representations. Notably, this line of thinking, which may seem more contemporary because of its appeal to neural systems, is also rooted in Jamesian thinking. In his chapter on attention, James speculated that attention operates at the neural level by activating neurons that receive sensory inputs pertaining to the attended stimulus in advance of their preferred stimulus. He wrote:

> *"The natural way of conceiving all this is under the symbolic form of a brain-cell played upon from two directions. Whilst the object excites it from without, other brain-cells, or perhaps spiritual forces, arouse it from within. The latter influence is the 'adaptation of attention.' The*

> *plenary energy of the brain-cell demands the co-operation of both factors: not when merely present, but when both present and attended to, is the object fully perceived."*

Thus, the contemporary definitions of attention just described cover many of the qualities described by James in 1890. This correspondence suggests that 19th-Century thinking about attention has guided more than a hundred years of research, and that even though with decades of ever-increasing effort we may have achieved better understanding of attention's neural mechanisms through refined research methods and improved technology, our theoretical understanding of attention nevertheless remains unchanged at some fundamental level. That is partially true, but contemporary thinking about attention is more nuanced.

Increasingly, psychologists and cognitive neuroscientists are espousing the argument that "attention" is a folk psychological concept that in fact might cover multiple distinct cognitive and neural phenomena, and using one term to refer to all of these phenomena is scientifically imprecise.

One strong form of this argument is that in philosophical terms, attention is not a natural kind, in the sense that it may not reflect a pattern or grouping of physical structures and processes that genuinely occurs in the natural world without reference to human actions or interests (Bird and Tobin, 2018). In other words, if attention is not a natural kind, we would not be able to find neural correlates of attention per se (Brigard, 2012). Attention would not exist independently of psychological theory. Instead, for example, it may be possible that what we consider to be attention might really be a mode of sensory processing that is distinct within each sensory domain – in that case attention would not be a common neural phenomenon that can be identified in both listening attentively and looking attentively. Therefore, it would be inaccurate to classify any one thing as “attention.” This is a strong argument because it disputes the existence of attention as a real entity.

A weaker (in the philosophical sense) and more widely accepted notion is simply that “attention” refers to numerous distinct cognitive and neural processes that may not all be co-occurrent. Some proponents of this idea suggest that a number of the neural processes and cognitive functions contained inside the “attention” label should be split off into their own categories, thus refining the definition of attention without jettisoning it altogether. Furthermore, the term “attention” has multiple usages in everyday language (Brigard, 2012), some of which are broad enough in what they refer to that it is unlikely those usages of the term map onto a single psychological process (Luck and Vecera, 2002).

Given the numerous arguments against the conventional classification of attention, why should we continue to use the term “attention” in cognitive neuroscience and psychology? Why should we not instead be focusing our efforts on delineating and substantiating the neural processes that constitute “attention,” and thereby make scientific progress on more solid theoretical grounds?

One reason is that although we are refining our understanding of the various neural processes that unfold as organisms perform attentive behaviors, there is no consensus about how to subdivide these various “attentional” processes into distinct constituent categories that are closer to natural kinds in a strict sense. In keeping with the weak argument that “attention” encapsulates numerous distinct cognitive phenomena, it would indeed be useful to reclassify previously-“attentional” phenomena that are better situated within a different category, so that the label “attention” can be reserved for a more unitary phenomenon and its scientific pursuit can be more precise. However, performing even this kind of weak response would require a firm understanding of why the phenomena to be reclassified should be excluded from the “attention” label, and this kind of understanding has not yet been achieved. A response to the strong argument that attention is not a real entity requires overcoming an even greater obstacle: In order to divide attention into its constituent natural kinds, those kinds would have to be categorically distinct from one another, a stringent criterion that would require clear demarcations between categories such that no smooth transition between categories or overlap between categories would be permitted (Bird and Tobin, 2018).

Another reason that we continue to study attention despite criticism of the term is that even if attention is not a natural kind, it is still a useful construct. Investigations into the neural mechanisms of attention are still yielding new information that is contributing to a firmer understanding of how the nervous system performs attentive behaviors, and no paradoxical results have yet arisen that would force the progression of theory towards new foundational definitions. Instead, often times the contrary seems to be true: The findings of many lines of inquiry have indicated that there are neural processes that subserve multiple kinds of attentive behavior, strengthening the notion that attention is a unitary phenomenon. Some of this evidence will be described in the following sections of this paper.

A third reason that the use of the term “attention” is still valuable in research settings is that the folk psychological nature of attention, the thing makes the concept of attention relatable to “everyone,” is itself motivational. The overarching mission of cognitive neuroscience is to uncover the neural mechanisms that underlie human mental experience, and so if attention is a key contributor to our mental experience, it merits investigation within our scientific discipline. This last argument sidesteps the criticism that “attention” may not refer to a natural kind by admitting that it is precisely because the definition of attention relates to human actions and interests that it should be investigated within the scientific field of cognitive neuroscience.

With the overarching mission of cognitive neuroscience firmly in view, the goal of the present chapter is to discuss the latest understanding of how the brain performs object-based visual attention, specifically in reference to natural objects. As I will argue in the following sections, object-based attention is a form of attention that is essential for effective behavior in natural environments. In other words, much of our quotidian mental experience involves not just attention, but object-based attention. I will start from the concept of attention, rooted in folk psychology and James-like thinking but nuanced with appreciation for more contemporary criticisms about the heterogeneity of processes under the “attention” label. Then I will move toward a review of object-recognition research, covering early models of how the visual system can recognize objects and neural evidence that refines and adds new details to these theories. Then, I will define object-based attention, review research on this topic, and draw upon findings from the object-recognition work to describe how object-recognition processes are linked to object-based attention. I will explain how object representations in the brain can act as substrates for the kinds of object-based attention that we commonly deploy in our everyday lives, and I will propose new research questions that build from this synthesis and can advance us toward a better understanding of the neural activity that generates and supports human experience. Finally, I will merge my speculative framework for object-based attention with a more formalized and well-established theoretical model of attention's neural mechanisms, the Specificity of Control (SPoC) Model, introduce a testable hypothesis that arises from this synthesis, and lay the groundwork for understanding the theoretical significance of empirical work along that line.

How is Attention Operationalized and Studied?

The concept of attention as described in the previous section suggests that attention is an essential component of all imaginable goal-oriented behaviors. It is hard to conceive of any behavior that does not require the selection of certain sensory or cognitive, internal or external information to the exclusion of other possibly targetable information, or the "focalization [and] concentration of consciousness," as William James put it. Even behaviors that are often associated with mind wandering and neglect for the task at hand, such as the commonly experienced phenomenon of driving a familiar road and arriving at one's destination without any recollection of the details of the trip (because one's mind was engaged in unrelated trains of thought), involve attention. We might say that task-oriented attention was absent or diminished in various aspects of the act of driving, but process-oriented attention was certainly involved in maintaining and following internal trains of thought (Luck and Vecera, 2002).

The fact that attention seems impossible to disentangle from waking behavior, and moreover the fact that attention is required to execute any task or follow any instructions, makes it seem that attention is not a suitable phenomenon for scientific study by experimental manipulation. Or at the very least, because attention is inherent in all behavior, attention as such is too broad to isolate and make any definitive empirical claims about.

However, researchers have devised ways to study attention despite these intrinsic difficulties. The key to studying attention empirically is to design experimental paradigms that necessitate attention's formally defined qualities. Attention can be operationalized by making empirical measurements that hinge upon its core qualities: the ability to elaborate or enhance information processing with the purpose of improving the efficacy, accuracy, or speed of behavioral responses. For example, a task that requires the discrimination of visual stimuli that are extremely difficult to distinguish, but can be discriminated under some behavioral conditions, implies that attention to the stimuli is operating more strongly or only operating in the conditions in which the stimuli can be discriminated, because one of attention's defining qualities is its ability to enhance sensory information processing. Thus, it could be said that attention was manipulated across the conditions of the task. Upon this logical foundation, attention becomes a natural phenomenon that can be studied directly. Its effects, its dynamics, and its underlying neural mechanisms are all available to empirical observation once it is operationalized.

Luck and Vecera group attention-manipulating experimental paradigms into four categories: cuing paradigms, search paradigms, filtering paradigms, and dual-task paradigms (Luck and Vecera, 2002). Each of these paradigms operationalizes attention by way of one or more of the defining qualities described previously.

Cuing paradigms involve instructing participants that they should guide their attention toward a specific target source of information. For example, in a spatial attention context, a cuing paradigm might involve instructing a participant to attend to either the left or right side of their visual field in anticipation of a target stimulus. These shifts of attention can be overt, following shifts of the eyes, or covert, in which the focus of attention is shifted without any concomitant eye movement. Using a cuing paradigm, attention’s effects on information processing are observable by contrasting responses to the cued and uncued sources of information (e.g., Posner, 1980). A crucial aspect of the cuing paradigm is that the cue instructs the participant to attend to a specific source of information, whether it be a region of visual space, an audio channel, a somatosensory region of the body or some other information source. This directedness necessitates the selective aspect of attention. Cuing paradigms are thus distinguished from alerting paradigms, which are similar in that they generally precede a target event with a warning signal, but without the strict requirement that the warning signal specifically direct a participant’s attention to a target information source to the exclusion of other potential information sources. Alerting paradigms are used to study general arousal and alerting systems, which are sometimes considered to be aspects of attention, especially in vernacular usage of term “attention” (Petersen and Posner, 2012). But the more common definition of attention in cognitive neuroscience excludes this phenomenon because it is weakly- or non-selective.

Search paradigms require participants to localize a target stimulus within an array of multiple nontarget stimuli. These paradigms engage a recognizable form of attention because they mimic natural, ecologically valid conditions of cluttered visual environments, albeit in an extremely simplified form. Natural environments are often crowded with abundant and highly variable visual information. Any given scene we encounter on a daily basis can include numerous moving or static objects arranged in countless ways that can be seen from multiple perspectives under different lighting conditions. In these natural environments ample with potential targets we often need to find specific objects in order to accomplish everyday tasks. Researchers bring a simplified, experimentally controllable version of this mundane – but complex – behavior into the laboratory in the form of a search paradigm. In a typical conjunction search paradigm, for example, the task might be to identify a lone green letter T from amongst a randomly arranged group of blue Ts and green Ls. Using classical designs such as this, researchers have found that search difficulty depends on the factors such as the number of items presented, the number of variable feature dimensions, the similarity of the target to the distractors, and the similarity of the distractors to each other (Duncan and Humphreys, 1989). However, recently attention researchers have found that characterizations of attention and search behavior made under classical search paradigms do not translate to more naturalistic forms of search (Peelen and Kastner, 2014). This issue will be discussed at length in subsequent sections.

Filtering paradigms invoke behavior that is associated with attention's ability to select targeted information to the exclusion of irrelevant or distracting information. These studies usually blend together target and distractor information at one source, so that performing the task and responding to target information requires suppressing interference from distractor information. For example, in a dichotomous listening task, two streams of speech may be presented at once, and only one stream of speech must be attended. In this case, speech is the singular information source, because distracting and target information are both included within that source.

Dual-task paradigms require participants to perform two independent tasks simultaneously, allowing researchers to observe how competition between the two tasks affects task performance. These studies evoke common real-world situations in which multiple tasks are vying for our attention at the same time, or when we are trying to multitask. Dual-task paradigms thus capture an aspect of attention that is not engaged in any of the previous three paradigm types.

Cuing, search, filtering, and dual-task paradigms are the canonical experimental designs that researchers use to isolate and manipulate attention, so that attention can be approached as an empirical entity. Most attention studies that are encountered in cognitive neuroscience and psychology research can be categorized into one of these paradigm types, but often a particular study design will incorporate elements of multiple of these types of paradigms. For example, an experiment may use instructional cues to inform participants about an upcoming target stimulus to be identified, following a cuing design, but the target may then also be embedded within noise or have other distracting features to be ignored, as in a filtering paradigm.

Regardless of the design details of an attention experiment, or which of the four types of paradigms an experiment could most easily be classified as, all attention experiments share the common fact that they operationalize attention by leveraging attention’s postulated psychological traits into measures of behavioral performance. For researchers who are skeptical of the existence of attention as a real entity that should be scientifically investigated as such, the logic underlying this strategy might seem circular and unsound: Attention is defined as possessing certain qualities, and these qualities are then utilized to yield observable measurements that are taken as evidence of the details of attention’s operation, and by implication, as evidence of attention’s very existence. If it were true that the conventional notion of attention did not accurately reflect the true cognitive and neural substrates of attention-like phenomena, we might expect that as researchers continue to examine attention with ever finer methodological and technological precision, and test new predictions born from previous findings, inexplicable results would emerge to challenge the standard model of attention. No crisis like this has occurred yet, which is not to say that it never will, but the preponderance of available evidence makes a strong case that the psychological qualities of attention such as the ones explicated by James in 1890 have neural substrates that can be identified. The next section will review important work that has begun to reveal these neural substrates.

Attention is Linked to Changes in Neural Activity

Operationalizing attention with designs involving cuing, search, filtering, or dual-tasks allows attention to be examined empirically. Cognitive neuroscientists aim to uncover the neural mechanisms of attention, and so by manipulating attention under the operational conditions described in the previous section, and concurrently recording neural data, researchers can link patterns of neural activity to the cognitive attentional processes being manipulated. In this section I will review findings from ERP, cell electrophysiology, and neuroimaging studies that have revealed how attention may be instantiated at the neural level. Then, I will review theories of how a top-down attentional control system might be influencing sensory neural activity to give rise to attention's effects.

Evidence from ERP studies

The first investigations of attention's neural correlates in humans used the electroencephalography (EEG) method. In EEG studies, an array of electrodes is affixed to the scalp to record a time series of voltage values that reflect summated electrical activity from neurons in the brain. These voltages do not reflect action potentials directly, but instead are attributed to post-synaptic potentials, primarily in cortical pyramidal neurons (Luck, 2014). EEG data taken from many trials from one condition of one experiment can be aligned to a common reference time point, such as the time at which a target stimulus onsets on each trial. Aligned data can then be averaged over trials, reducing inter-trial variability attributable to experimental noise and fluctuations in mental state, to yield event-related potential (ERP) waveforms that reflect neural processes that are common to the experimental condition in question.

The comparison of ERP waveforms between conditions is a valuable tool for examining the neural correlates of cognitive phenomena such as attention. Components of ERP waveforms cannot be precisely localized to brain areas for numerous reasons, including the smearing of electrical fields that occurs as the fields emanate from dipole sources in cortex through tissue layers and skull before they are captured by recording electrodes on the scalp. However, the temporal precision of EEG is extremely high, and with a high enough sampling rate one can track electrical potential changes arbitrarily closely to their origin in post-synaptic potential changes. This allows for fine-grained analysis of the time courses of various neural processes and how these time courses may differ between experimental conditions. Furthermore, amplitudes of ERP components can be compared between conditions, in order to make distinctions between various conditions in terms of the degree of neural activity at a certain processing stage that accompanies each condition. Researchers have made use of both of these advantages of the ERP method to address questions about the neural correlates of attention.

At the time that the initial electrophysiological studies of attention were being conducted, there was a significant debate over whether attention operated over early sensory input processes or only at later post-perceptual stages. In the early selection theory, attention was thought to exert effects on early neural responses to incoming sensory signals, enhancing the representations or influence of attended stimuli in sensory systems so that those representations could be more efficiently integrated into later task-execution computations. According to the late selection theory, stimulus processing and identification happened in parallel and with equal efficacy across the whole visual field, and cognitive capacity limitations only occurred later in decision making stages. By this theory, attention was thought to operate over later-stage decision making or memory processes (Luck et al., 1994).

Behavioral results favored the early selection theory. For example, Luck and colleagues investigated whether the behavioral effects of attention on reaction time between valid and invalidly cued targets could be attributable to memory or decision-related factors (Luck et al., 1994). The researchers ruled out several interpretations of previous studies' findings that had been proposed in light of the late selection model. In one experiment, Luck and colleagues tested whether reduced sensitivity to target stimuli at uncued locations could be due to memory limitations, as late selection advocates argued, by adjusting the cue period duration. The researchers found that cue duration did not change the observed attention effects. In a second experiment, Luck and colleagues tested whether cue validity effects could be attributable to decision processes and not sensory processes by inserting an abrupt onset post-cue stimulus that would be expected to override cued attentional state and hinder the transfer of sensory information to decision making stages under the late selection model. The researchers found that cuing effects were not diminished by salient post-cues, strengthening the early selection interpretation of previous findings.

Because of its high temporal resolution, ERP was ideally situated to address the early/late selection debate neuroscientifically. ERP studies of visual attention found that attention operates at an early stage of sensory processing. Numerous studies suggest that cuing attention toward a location in space affects early sensory processing of target stimuli when they appear in the attended region of space (Mangun, 1995; Hillyard et al., 1998). In these studies, the amplitudes of ERP components found most strongly over occipital scalp in the range of 50 – 250 msec after target onset differed depending on whether the target appeared at the attended location or at an unattended location. These components, known as P1 (80 – 130 msec) and N1 (170 – 210 msec), are thought to reflect neural activity in early stages of visual processing. Thus, when the amplitudes of these components are shown to increase with attention to the ERP-eliciting visual stimulus, it is taken as evidence that attention voluntarily allocated by the instructional cue modulates the processing of stimulus inputs in early visual cortex. This finding fits with the early selection theory's proposal that attention enhances sensory perceptibility of stimuli by directly increasing the early neural response to those stimuli, or by increasing the signal to noise ratio of the neural representation of those stimuli.

To clarify the picture of how attention alters early sensory processing, Luck and colleagues investigated the enhancement of neural responses to attended stimuli and the suppression of neural responses to unattended stimuli, showing that these two sensory response modulations are attributable to independent mechanisms (Luck et al., 1994). The researchers were able to dissociate the effects of cued attention on the P1 and N1 ERP components: P1 amplitudes were increased to attended targets, and N1 amplitudes were decreased to unattended targets. This finding suggests that multiple attentional mechanisms operate at early sensory processing stages, both to increase the neural response to attended stimuli, and to decrease the response to unattended stimuli. This dissociation reflects attention's psychological property of selecting targeted information to the exclusion of other potentially targetable information, and adds the new detail that exclusion of task-irrelevant information does not happen passively by relation to enhanced task-relevant information, but rather that task-irrelevant sensory inputs are subjected to an actively suppressive influence.

These findings exemplify how the body of ERP studies on attention supports an early selection model. By the early selection model, attention exerts modulatory control over sensory representations as they enter sensory cortex and not only at higher-level decision making and task execution stages of cognition. ERP studies therefore also support a gain control model of attention, by which attention's cognitive and behavioral effects result from a strengthened neural representation of task-relevant stimuli (Hillyard et al., 1998). These studies utilized the ERP method's temporal precision to address questions about the timing of attention's modulatory influences on neural activity. Another benefit of the ERP method is that it is non-invasive, allowing direct measures of neural activity to be taken from human participants. However, the non-invasiveness of EEG recording is also a limitation. The ERP method is limited in its ability to localize effects to brain areas, and in its ability to more closely examine the effects of attention at the level of individual neurons, where neural computations take place.

Evidence from cell electrophysiology studies

Because EEG electrodes record electrical activity at the scalp, they reflect summated activity of millions of neurons over a widespread cortical patch. To gain a better understanding of attention's effects on neural activity at the level of individual cells, researchers perform electrophysiological studies on brain cells of animal models. Numerous cell electrophysiological studies have revealed details of how attention modulates the magnitudes of neural responses to attended and unattended stimuli at the level of single neurons, corroborating findings from ERP studies, and suggesting a more fine-grained picture of how many individual neurons' modulated responses result in ERP component amplitude changes.

Moran and Desimone recorded from single cells in rhesus monkey visual cortex, and found that in order to enhance processing of attended visual stimuli and suppress processing of irrelevant stimuli, attention modified visual neurons' receptive field properties (Moran and Desimone, 1985). For their recordings, the researchers identified V4 neurons that had receptive fields large enough to encompass two different stimuli presented on a display simultaneously. At the V4 stage of visual processing, neurons are responsive over large spatial receptive fields but are selective for other visual features such as orientation and color. Therefore, one of the displayed stimuli could be designed to be effective in driving action potential activity in the recorded neuron, and one of the stimuli was not. This differential response elicitation profile of the two visual stimuli allowed Moran and Desimone to observe how attention modulated the action potential trains from the recorded neuron, purely as a function of which stimulus the monkey was attending. The monkey was trained to covertly attend to one stimulus for a block of trials and then covertly attend to the other stimulus for a subsequent block of trials, maintaining its fixation the entire time. This way, visual input was identical between blocks. The researchers found that when the monkey attended to the effective stimulus, the cell responded well, but when the monkey attended to the ineffective stimulus, the cell's response was greatly reduced, even though the effective stimulus was still inside the receptive field of the recorded neuron. The contrast in the cell's response between attention conditions suggested that attention can attenuate the response of neurons that represent unattended, task-irrelevant information. Furthermore, Moran and Desimone observed that there was no effect of attention on the recorded spike rate when one of the two stimuli was located outside of the recorded cell's receptive field. The researchers interpreted this result to mean that attention serves to bias competitive interactions between stimuli contained within a cell's receptive field, so that the cell propagates information about the attended stimulus and not about task-irrelevant stimuli.

The idea that attention can modulate neural responses to different stimuli within a single receptive field received more support from a study by Luck, Chelazzi, Hillyard, and Desimone (Luck et al., 1997). In their study, Luck and colleagues recorded from neurons of a macaque monkey in visual areas V1, V2, and V4 as the monkey performed a covert spatial attention task, shifting its attention between stimuli at two locations. For recordings made in areas V2 and V4, the researchers found that when the two stimuli were both within the recorded neuron's receptive field, the neuron's response varied according to which stimulus was attended, just as in the study by Moran and Desimone. Furthermore, baseline firing rates of the recorded cells before stimulus onset were found to be increased when the monkey's attention was directed into the receptive field of recorded cell. This baseline firing rate increase can be construed as preparatory biasing of upcoming sensory inputs from the attended location. No baseline firing rate increases or stimulus evoked activity modulations were observed in area V1, in line with previous studies that suggested that attention's effects on visual neural activity occurred in extrastriate cortex.

Human ERP studies showed that attention amplifies the neural responses to attended stimuli. Single unit recordings showed similar response enhancement in cells that were receptive to input from attended stimuli, suggesting a neuron-level basis for the ERP findings obtained from human participants. But other forms of neural activity modulation could also correlate with attention. Neural response amplification findings support gain control models of attention, however, gain control models cannot fully account for the important role that attention plays in shaping the flow of information through sensory processing areas and toward task-oriented behavioral execution areas. The single unit recording studies of Moran and Desimone and Luck and colleagues suggest that attention may be performing more than just an amplification on stimulus response at the individual neuron level, but may also be biasing attended stimulus representation relative to the representations of competing stimuli. Building upon gain control results, numerous cell electrophysiological studies suggest ways that attention alters the functional connectivity of brain areas in sensory processing hierarchies.

Fries, Reynolds, Rorie, and Desimone recorded spikes and local field potentials from clusters of multiple V4 neurons in a macaque monkey performing a covert spatial attention task, and found increased gamma frequency synchronization across neurons responsive to attended stimuli, compared to the coherence across neurons responsive to unattended stimuli (Fries et al., 2001). Fries and colleagues postulated that gamma frequency coherence among neurons reflects an open channel for neural communication (Fries, 2005), so the observation that gamma band synchrony among cells responsive to an attended stimulus suggests that one of the neural mechanisms of attention is regulating the functional communication pathways in cortical networks.

Thus, regulating the functional connectivity between cortical areas may be one of the neural mechanisms of attention. With attention controlling the flow of sensory information through cortical processing stages, task-relevant information can be processed preferentially or propagated efficiently toward higher order areas. Exactly how functional connectivity between cortical areas is controlled is unknown, however, there is some evidence that subcortical structures in the thalamus, such as the pulvinar nucleus, may be involved (Saalmann et al., 2012).

Although recording activity from single cells in attentive animals may not translate entirely to understanding the neural mechanisms of attention in humans, the findings just described were taken from primate animal models, and there is reason to believe that the operation of primate visual systems is highly conserved. Thus, cell electrophysiology in non-human primate models is an important method for elucidating the underlying neural mechanisms of attention in humans, when the dynamics of individual neurons in the human brain are not amenable to observation.

Evidence from neuroimaging

Human neuroimaging is a useful method for localizing brain activity to specific anatomical areas. Unlike EEG, which is anatomically imprecise but high-resolution in time, neuroimaging methods allow for identification of active brain areas down to cubic millimeter resolution, but temporally smear activity over seconds, precluding any observation of how localized neural activity evolves at time scales most relevant to neural computation. Like EEG, the major advantage of neuroimaging is that it can be performed in human participants, in non-invasive ways involving functional magnetic resonance imaging (fMRI) or only mildly invasive ways involving positron emission tomography (PET).

Despite its limitations, neuroimaging is well-suited to studies of attention's neural correlates. Neuroimaging's poor temporal resolution is less problematic within experimental designs that invoke sustained, top-down activity, such as the anticipatory attention elicited by cuing paradigms. Because findings from cell electrophysiology studies have shown that attention increases the baseline firing rate of neurons responsive to attended stimulus attributes (Moran and Desimone, 1985; Luck et al., 1997), overall brain activity in regions containing those responsive neurons would be predicted to increase during a period of sustained top-down attention. Furthermore, the spatial resolution of neuroimaging methods is sufficient for analysis of brain activity in regions of interest and at scales that are hypothesized to be relevant for attention's operation. Visual areas selectively responsive to different visual information can be circumscribed and observed within the spatial limits of fMRI (Grill-Spector and Malach, 2004).

Corbetta and colleagues used PET to show that activity in areas of the visual system specialized for different types of visual information was selectively modulated by attention (Corbetta et al., 1990). The researchers used a PET method that measures changes in cerebral blood flow. Therefore, the signal they acquired was used a proxy for underlying neural activity, and changes detected in the PET scans were considered to represent concomitant changes in neural activity across the entire cortical area where the signal was detected. The task that the researchers designed required that participants engage top-down attention toward different visual feature dimensions – shape, color, and velocity of stimuli – so that neural activity could be assessed in the visual areas that have been identified as selectively responsive to those features. The researchers' finding that brain activity increased in visual areas that encode attended feature dimensions matches findings from fMRI studies.

Functional MRI measures blood oxygen level dependent (BOLD) activity. Increased blood flow to a brain area is thought to reflect greater metabolic activity due to greater neural activity in that area, and so the BOLD signal is used as a proxy for neural activity (Logothetis and Wandell, 2004). Thus inference about neural activity on the basis of fMRI is similar to that of PET, but fMRI research is generally less invasive than PET and so it is more widely used in cognitive neuroscience.

Numerous fMRI studies have examined the neural correlates of top-down attention on sensory activity. Tootell and colleagues, using a covert spatial attention design, showed that BOLD activity was increased over visual areas that receive input from the attended visual field quadrant (Tootell et al., 1998). The degree of signal change was found to increase along the hierarchy of visual processing areas, which may reflect the enlargement of receptive fields with successive visual areas, and corresponds to previous findings of attentional modulation primarily in extrastriate areas. Notably, though, Tootell and colleagues did observe modulated BOLD activity in primary visual cortex. A similar finding was later reported by Martínez and colleagues (Martínez et al., 1999). However, Martínez and colleagues hypothesized that the modulation of BOLD activity in striate cortex reflected re-entrant processing that may occur as feedback from higher visual areas, because they did not observe any ERP changes in the time ranges that correspond to the initial sweep of sensory input through V1.

Kastner and colleagues tested a prediction in humans based on the finding from single-unit recordings in monkeys that when multiple stimuli are simultaneously present within a neuron's receptive field, they mutually compete for representation by the neuron, and attention can influence the neuron's responsivity (Moran and Desimone, 1985). Kastner and colleagues conducted an experiment in which participants' spatial attention was directed to a region of space containing four different visual stimuli, with instructions to attend to only one stimulus. The researchers observed that when the target stimulus was covertly attended, suppressive influence attributable to the other simultaneously presented stimuli was reduced, in line with the finding from monkey single-unit work (Kastner et al., 1998).

The studies reviewed here show how neuroimaging has been used to uncover attention's neural correlates in humans. The finding that top-down anticipatory attention correlates with an increase in activity in visual areas that selectively respond to the attended visual information resembles findings from single-unit studies that show an increase in baseline firing rates of individual neurons responsive to the attended information. Converging evidence from these two methods strengthens the idea that in part, attention acts as a gain control mechanism at the level of neural computation over sensory input. Despite its limitations, neuroimaging studies extend findings from cell electrophysiology: The human visual system responds in a way that is predicted from single-unit recordings in animal models, and baseline activity increases can be observed over an entire cortical area when that area is functionally selective for an attended source of visual information.

In addition to its applicability in humans, another advantage of neuroimaging not afforded by single-unit recording is that activity in the whole brain can be observed simultaneously. Thus, neuroimaging data can be analyzed to look for brain areas functionally linked to the baseline increases in sensory areas that attention induces. For example, the degree of attention-related change in activity in the lateral geniculate nucleus (LGN), the thalamic nucleus that receives afferent optic nerve signals and relays these signals to visual cortex, has been found to correlate with that of visual cortex, suggesting that attention might exert a gating influence subcortically (O'Connor et al., 2002). Concomitant activity has also been hypothesized to exist in higher-order brain areas associated with cognitive control, because top-down attentional modulation of sensory processing theoretically is implemented from a directive source. The next section will review human neuroimaging studies that have sought to reveal sources of attentional sensory modulation that constitute a theoretical attention control system.

Evidence of a top-down control system

Three ideas motivate the hypothesis that attentional modulation of sensory activity is a result of activity from a higher-order attention control source system. First, without any top-down or feedback influence, inputs to early sensory areas are unbiased and contain all sensory information available to subsequent analysis. Abstract representations such as features and meaning are developed as sensory input progresses through increasingly higher-order stages of processing. In that sense, the earliest sensory representation can be thought of as a blank slate or a pluripotent map of the sensory environment. Therefore, attentional modulation of sensory input, which has been shown to selectively emphasize representations of some sensory information to the exclusion of other information, cannot occur without top-down influence.

Second, higher-order brain areas have been associated with cognitive control, broadly defined as the cognitive function that enables the control of behavior and internal cognitive operations that support a goal or task. Within this conceptual framework, endogenous attention, the form of attention engendered by voluntarily orienting attention toward a source of task-relevant information, can be linked to cognitive control. Thus, the orienting of endogenous attention is thought to originate from higher-order brain areas.

Third, in clinical studies, patients with lesions in frontal and parietal areas often show specific deficits in their ability to attend. Corbetta and Shulman have reviewed how damage in the right temporoparietal cortical junction (TPJ) often leads to spatial neglect of the contralateral hemifield – a syndrome that has been construed as a specific deficit in the ability to direct attention to the contralesional visual field or contralesional half of objects (Corbetta and Shulman, 2002). Neuropsychology findings thus provide causal evidence that areas in the parietal lobe are involved in the orienting of attention.

On the basis of these three motivational arguments for the existence of a higher-order attention control system, researchers have sought to uncover the brain areas that may constitute such a system, what the respective role of each area is, and how their cooperation mechanistically modulates sensory processing. Crucially, these efforts have separated brain activity attributable to attentional orientation from attentionally modulated sensory activity, in order to examine the control processes themselves.

Numerous early studies identified a collection of areas that are involved in the orienting of attention, without experimentally evaluating what the differential role of each area might be. Using PET, Nobre and colleagues observed that areas including the right anterior cingulate gyrus, the intraparietal sulcus of right posterior parietal cortex, and the medial and lateral premotor cortices were involved in both exogenous and endogenously cued covert spatial attention (Nobre et al., 1997). The researchers designed a covert attention task with peripheral cues. In one condition, participants were instructed that the peripheral cue appears on the side where the subsequent target will occur. This condition engendered exogenous attention, because the peripheral cue was a salient stimulus that captured attention and caused its reflexive orientation toward the cued side. In a second condition, participants were instructed that the target would appear on the opposite side of where the peripheral cue appears. This condition was designed to engender endogenous attention, because participants would have to orient attention voluntarily in accordance with the task instructions, rather than rely on the reflexive orienting that occurs with the peripheral cue. In comparing PET results between the endogenous and exogenous attention tasks, Nobre and colleagues observed that the activity in frontal and parietal areas seemed to be equivalent across conditions.

From the results of Nobre and colleagues, it is unclear whether identical network activity generates both endogenous and exogenous shifts of attention. These two forms of attentional orienting seem conceptually distinct, and therefore it is unlikely that they are attributable to identical network mechanisms. Corbetta and colleagues used event-related fMRI to examine attention-related parietal activation, such as that observed by Nobre's group, to closely inspect whether different regions within parietal cortex were differentially active depending on whether attention was voluntarily oriented in an endogenous manner or reflexively oriented exogenously (Corbetta et al., 2000). The researchers found that the intraparietal sulcus was preferentially involved in endogenous attention, and the right temporoparietal junction was more involved during exogenous orientation.

Hopfinger, Buonocore, and Mangun used event-related fMRI and a cued attention task to dissociate cue-related neural activity from target-related neural activity (Hopfinger et al., 2000). Cue-related activity was thought to reflect voluntary orienting processes, and localized to a network of cortical areas including superior frontal, inferior parietal and superior temporal brain regions. The researchers also found greater BOLD activations in extrastriate areas evoked by the cue, in line with previous findings that endogenous attention increases baseline activity in sensory areas that encode attended information, and strengthening the interpretation that the network of active areas in frontal, parietal, and temporal cortex influenced the receptivity of visual areas to attended inputs.

Beyond just identifying the brain areas that constitute the attention control network, understanding the computational or mechanistic contributions of each of the areas is a key goal of research into the neural mechanisms of attention. Several groups have examined the role of individual areas in this network and the interaction between components of the network in order to better understand the respective contributions of each area.

For example, Szczepanski and colleagues used diffusion-weighted and functional MRI to investigate how different combinations of the constituent areas of the dorsal frontoparietal attention network differentially contribute to attentional control (Szczepanski et al., 2013). The researchers found that different functional pathways between areas in the dorsal attention network were linked to different spatial representations of attentional priorities. Egocentric spatial reference frames, in which positions in space are derived relative to the observer's viewpoint, are different from allocentric reference frames, in which spatial coordinates are defined relative to an external object. Szczepanski and colleagues found that spatial attention operating within these two reference frames depended on different functional pathways between nodes of the dorsal attention network. Furthermore, the researchers found that these different functional pathways may be subserved by different anatomical pathways. The researchers compared activity between two groups of nodes in the dorsal network: the frontal eye field (FEF) and intraparietal sulcus (IPS) group, and the supplementary eye field (SEF) and superior parietal lobule (SPL) group. The researchers concluded that unlike FEF and IPS, which preferentially represented space egocentrically, SEF and SPL could flexibly represent space either egocentrically or allocentrically, depending on the task demands. Thus, these findings lead to a better understanding of the respective roles of the different areas in the dorsal attention network.

Popov, Kastner, and Jensen recorded magnetoencephalography (MEG) from human participants performing a visuospatial attention task, and observed that of all the areas in the dorsal attention network, the right FEF was preferentially involved in the control of gamma-band frequency activity in visual cortex via alpha-band activity (Popov et al., 2017). The gamma band, defined as oscillatory neural activity in the range of 40 – 100 Hz, is thought to reflect local neural activity associated with the feedforward flow of sensory information, whereas the alpha band, defined as 8 – 13 Hz oscillatory activity, reflects feedback from control regions onto sensory areas. Thus, the researchers concluded that within the conceptual framework that distinct bands of oscillatory activity reflect separate channels for information transmission, the right FEF plays a critical role in relaying attentional control signals generated in the entire dorsal attention network to the sensory sites of attentional modulation. Whether this finding reflects a unique role for the right FEF is debatable, because the right FEF is not known to possess a retinotopic map of the complete visual field (Silver and Kastner, 2009). Instead, the finding of Popov and colleagues may only partially reflect a more complete bilateral FEF mechanism due to methodological limitations.

In summary, a dorsal frontoparietal attention control network has been identified using human neuroimaging. The brain areas that constitute the dorsal attention network include IPS, SPL, FEF, and SEF of both hemispheres (Szczepanski et al., 2013). This network has been implicated in endogenous attentional orienting, and its activity has been linked to modulations of neural activity in sensory areas (Liu et al., 2016; Popov et al., 2017), but future studies are necessary for better understanding of the precise mechanisms by which activity in the dorsal attention network modulates sensory representations of attended information.

One possible mechanism involves channeling control signals to sensory cortex through retinotopic maps of space. Both the IPS and the FEF exhibit retinotopically organized maps of visual space (Silver and Kastner, 2009). These spatial maps may be involved in directing covert and overt attention, and thus activity in these areas might give rise to the network activity observed during endogenous spatial attention (Scolari et al., 2015). Activity in this network has also been involved in feature-based attention and object-based attention, suggesting that visual feature locations can be identified and attended through the retinotopic maps encoded in IPS and FEF (Scolari et al., 2015).

The dorsal attention network is hypothesized to be the primary control network involved in endogenous attention. However, in controlling attention to execute goals and implement task-relevant behavior, the dorsal attention network does not operate in isolation. The dorsal attention network is thought to closely interact with a separate ventral frontoparietal attention network (Vossel et al., 2014). The ventral frontoparietal network, comprising the temporoparietal junction (TPJ) and ventral frontal cortex (VFC), has been identified as a source of exogenous orienting (Corbetta and Shulman, 2002), potentially acting on sensory cortex via connections to the dorsal network (Corbetta et al., 2008; Geng and Mangun, 2011). A commonly proposed mechanistic role of this ventral network in attentional reorienting is that it may act like a circuit breaker for the dorsal network, triggering shifts of attention when highly salient stimuli reach a critical threshold. However, there is debate over this claim, in part because there is little direct evidence that the network plays this circuit breaker role, and because other evidence suggests that it may instead primarily integrate contextual information into attentional orienting (Geng and Vossel, 2013). The ventral frontoparietal attention network is conventionally defined as a right hemisphere network because of clinical findings from hemispatial neglect syndrome patients (Corbetta and Shulman, 2002), however, neuroimaging experiments have revealed that corresponding left hemisphere areas may also play an important role in orienting attention on the basis of contextual information (DiQuattro and Geng, 2011; Geng and Vossel, 2013).

Indeed, the finding that a network of higher-order brain areas seems to control exogenous attention is surprising. A prediction that higher-order areas control exogenous attention does not follow directly from theory in the same way that endogenous attention was predicted to originate from sources in higher-order areas. Mutually competitive interactions mediated by inhibitory interneurons exist throughout the cortex, and such inhibitory architecture can be used as a mechanism to emphasize and strengthen highly salient inputs to the exclusion of less salient inputs (Coultrip et al., 1992). Thus, on the basis of these winner-take-all models, exogenous orienting of attention could be predicted to operate entirely from bottom-up mechanisms, without any involvement from higher-order orienting networks. Future work will be necessary to better understand the role of the ventral frontoparietal attention network, and integrate early interpretations of the network as a reorienting controller with more recent findings about its representations of contextual information.

How are Objects Represented in the Brain?

Higher-order attention networks have been shown to be involved in the orienting of attention. Neuroimaging studies in humans have revealed that dorsal and ventral frontoparietal networks are active when people orient their attention toward task-relevant or salient sources of sensory information. These networks, particularly the dorsal attention network, are considered to be the source of directive signals that influence the responsivity and computational efficacy of selected sensory areas that represent the kind of sensory information necessary to efficiently perform a task or execute a goal. It is unknown exactly how functional connections from the dorsal attention network onto sensory sites lead to the attention-mediated changes in neural activity that have been observed using ERP, cell electrophysiology, and neuroimaging methods, and moreover, it is unclear exactly how control signals are channeled to sensory areas. Regarding the latter point, researchers have identified that the IPS and FEF contain topographic spatial maps (Silver and Kastner, 2009), suggesting that control signals generated by activity in the dorsal attention network are channeled through these areas in order to target selected visual sites (Scolari et al., 2015). In support of this idea, researchers have observed that of all the components of the dorsal attention network, the FEF is most clearly functionally linked to attention-related visual modulation (Popov et al., 2017).

But the number of anatomical connections necessary to support attention to every possible spatial location, spatial scale, visual feature, and object category undermines the possibility that the dorsal attention network operates via a one-to-one mapping of task demand to sensory modulation. Alternatively, attention may operate primarily over object representations, reducing the complexity of the signal transmission problem. The aim of the present manuscript is to describe how attention may operate in this manner, and in order to contextualize that topic for more detailed discussion later, in this section I will describe the current understanding of how visual objects are represented in the brain. First, I will describe early models of how the visual system recognizes objects, and provide an overview of the development of object recognition theory to its current state. Then, I will review work from cell electrophysiology and human cognitive neuroscience that examine how the brain instantiates models of object recognition.

Early models and behavior

In his book *Vision*, David Marr defined vision as “knowing what is where by looking” (Marr, 1982). Marr’s quip refers to an essential quality of vision – recognizing the objects that we see. In the vast majority of cases throughout our lives, when we look at something, we immediately know what it is. But importantly, Marr’s satirically facile definition of vision hints at a complex problem under the surface: What is an object? It seems trivial, because intuitively we know what objects are. We interact with the world primarily on the basis of the objects that fill our environments. Typical scenes are cluttered with many objects. In almost all cases, we recognize objects immediately from normal visual input, and we have little trouble recognizing objects even as our perspective on them, their relative size, their positions relative to other objects, and their lighting conditions change. We easily and naturally generalize our learned object categories to new exemplars that exhibit combinations of visual features that we have never encountered before and never consciously anticipated. As Marr implied, our easy everyday experience with objects belies the difficult and complex computational problems that our visual systems overcome to generate and sustain our object perceptions.

One major problem is reconciling selectivity with invariance: our visual system needs to be able to immediately recognize an object with a high degree of certainty, selecting one identity over all other possible identities, but it also needs to be able to perform this selective identification invariantly over a wide range of visual appearances (Gauthier and Tarr, 2016; Tong, 2018). Another major problem the visual system faces is extracting three-dimensional object information from two dimensional projections of light onto our retinas, when an infinite number of possible arrangements of entities in the real world could give rise to the single two dimensional retinal projection each eye receives (Tarr and Bülthoff, 1998).

Evidently, the visual system can overcome these computational challenges, and a key component of its solution, or the solution itself, is the formation and utilization of object representations. Thus understanding how objects are represented in the brain is fundamental to a cognitive neuroscience of vision. The first step toward reaching that understanding is defining what objects are.

We can define an object as a collection of visual features, none of which are individually necessary or sufficient for categorization of the collection as an object, but that together are perceived as a bound, discrete perceptual unit. Objects have affordances, meaning that our perception of an object incorporates the ways that we can interact with it, manipulate it, and use it. Objects often have verbal labels, but the binding of visual features into an object can still occur without invoking a verbal label. When an object has a verbal label, we say that it belongs to a category. We can make a distinction between an object category and subordinate members of a category. Objects are thought to reflect, imperfectly but closely, the natural category structure of the world, according to which object categories maximize the similarity of members of the category while also maximizing the separation between categories. In line with this idea, researchers have found behavioral evidence that people are faster at identifying objects at the basic category level than they are at the subordinate level, suggesting that more cognitive effort and neural computation is required to discriminate subordinate-level categories (Mervis and Rosch, 1981; Palmeri and Gauthier, 2004).

Research into the neural representation of objects seemingly can be differentiated according to two broad questions: How does the brain extract perceptual features and bind them together into a coherent, discrete object representation? And: What is the neural basis of object category classification? These two seemingly disparate processes involving objects are actually interconnected. Grill-Spector and Kanwisher conducted a behavioral experiment to test whether object segmentation, the process of identifying the location and extent of an object within a visual scene, occurs before object categorization. The researchers presented participants with object images for brief durations and followed by a visual mask, in order to limit the processing time that the images could bestow on the visual system, and then asked participants to report on the presence, category, or the subordinate-level identity of the presented object. If object segmentation occurs before categorization and subordinate-level identification in the visual processing hierarchy, participants should more easily report the presence of an object than the category of an object. Grill-Spector and Kanwisher found nearly identical response profiles for detection and categorization, with identification lagging behind, and interpreted their findings to mean that object segmentation – the binding of visual features into a discrete visual representation – occurs contemporaneously with object categorization (Grill-Spector and Kanwisher, 2005). Thus, the process of forming coherent object representations in the visual system may be identical with the process of categorizing objects, or at least, models of object recognition are constrained to those that do not require separate processes for object segmentation and categorization.

With the concept of an object defined along the lines described above, researchers have sought to model how the visual system extracts object information from visual input. Early models of how the brain represents and recognizes objects emphasized that the brain extracts three-dimensional structure from two-dimensional retinal projections. These models are broadly classified as structural description models (Tarr and Bülthoff, 1998). For example, Marr proposed a “2.5-D sketch” process, by which the brain constructs a three-dimensional model of an object in an allocentric reference frame from the two-dimensional retinal projections we receive in an egocentric reference frame (Marr, 1982). According to Marr’s model, the visual system builds from lines to contours to surfaces to three-dimensional object representations, to reconstruct the external source of input to our retinas. In a similar vein, Biederman developed a “recognition-by-components” or “geon” model, according to which every object is represented primarily as a combination of a small number of elementary three-dimensional shapes (Biederman, 1987). Support for structural description models has been gleaned from behavioral studies showing that object recognition can be primed by presenting participants with pictures of the target objects but from different perspectives (Biederman and Gerhardstein, 1993). These models are plausible, because they propose that by extracting a three-dimensional object structure from visual input, the visual system circumvents the need to hold a representation of every classifiable object from every viewpoint from which the object might observed. But criticism of both Marr's model and Biederman's model includes the assertion that neither model explains the actual computational processes underlying three-dimensional structure extraction, and the mystery of that process is the critical gap in our understanding of object recognition (Tong, 2018). Furthermore, Biederman's geon model cannot be easily applied to the recognition of natural objects, particularly when the geon deconstruction of two completely different objects is identical.

In contrast with structural description models, a different wave of object recognition study has been based on the idea that objects are represented as a collection of viewpoint-specific features. These models are classified as image-based models (Tarr and Bülthoff, 1998). Image-based models are inherently dubious because they suggest that the visual system does encode separate images for all or many of the possible viewpoints that every recognizable object can inhabit. But researchers in the late 1980s and 1990s suggested that the visual system only needs to encode a smaller number of viewpoint representations, and then can rely on mechanisms such as mental rotation and view interpolation to accommodate the reality that an object can be recognizable from an infinite number of perspectives (Tarr and Bülthoff, 1998).

Ultimately, neuroscientific studies that took place in the late 1990s and first two decades of the 21st Century have rendered moot the debate between proponents of the image-based and structural description models (Gauthier and Tarr, 2016). A new class of models – convolutional neural networks – has eclipsed previous theoretical accounts of object recognition. The next section will describe this newly dominant theoretical framework.

Convolutional neural network models

Artificial deep convolutional neural networks (CNNs) are currently the most plausible and accurate models of how the human visual system performs object recognition (Kriegeskorte, 2015; Gauthier and Tarr, 2016; Tong, 2018). Although CNNs were not invented with the explicit intention of replicating the brain's visual system, the design of CNNs is inspired by the biological architecture of the visual system, and resembles particularly closely the ventral visual pathway (Kriegeskorte, 2015). In a typical CNN architecture, interconnected layers of nodes perform simple computations, passing an incoming visual representation forward through successive layers of computation until it reaches an output layer. Activation of one out of an arbitrary number of nodes in the output layer signifies the class of object depicted in the visual information given to the CNN as input. In this kind of architecture, individual nodes are the basic computational units, and the overall computation is instantiated in the network interactions between the nodes. CNNs are defined as deep when they comprise many hidden layers – in current practice, typically at least five – between the input layer and the output layer. These networks are convolutional in the sense that the early layers have small spatial receptive fields, and because each node in these early layers performs a feature extraction over its receptive field. Hence, the cumulative effect of all nodes' operation is a convolved feature extraction over the input image space.

The CNN design is thought to resemble the way that biological neural networks perform computations over sensory input, but artificial neural networks generally do not incorporate many of the theorized operational mechanisms of biological neurons, such as dendritic computations, differential excitatory efficacy along the membrane, and the multiplicity of computational roles played by different types of neurons.

Nevertheless, remarkably, CNNs perform many high-level object classification tasks as well as humans in some respects. For example, Rajalingham and colleagues collected a large scale dataset of object recognition behavior from monkeys and humans and compared it to feedforward CNN performance, finding that CNNs, monkeys, and humans all shared a highly similar pattern of object-level confusion – showing a nearly identical tendency to mistakenly classify an object category as a different category – indicating similar underlying object classification mechanisms (Rajalingham et al., 2018). However, notably, Rajalingham and colleagues observed that the CNN did not perform similarly to monkeys or humans at the individual image level. That is, for each individual image tested, the CNN's classification performance did not correlate strongly with that of the monkeys or the humans. The researchers' finding highlighted a limitation in drawing an exact comparison between feedforward CNNs and primate object recognition networks. A strong claim that artificial neural networks replicate primate object recognition behavior would require a close correspondence between model and primate performance on all available metrics.

Despite the limitations observed by Rajalingham and colleagues, the correspondences in design and performance that have been observed between the human visual system and CNNs suggest that understanding how CNNs perform object classification is at least relevant to understanding how the human visual system represents and recognizes objects, if not fully explanatory. To that end, researchers have shown close correspondences between internal object recognition processes in CNNs and in biological visual systems. Khaligh-Razavi & Kriegeskorte, using representational similarity analysis (Kriegeskorte et al., 2008), showed that the way that early CNN layers represent visual inputs match the early stages in the biological visual hierarchy, whereas the representational geometry of later layers of CNNs more closely match that of inferotemporal cortex (Khaligh-Razavi and Kriegeskorte, 2014).

Many high-performing CNNs are exclusively feed-forward, meaning that once the network has been trained, the processing of representations at one layer is not influenced by the activity in any of the subsequent layers. However, the ventral visual pathway of the human brain contains many lateral and feedback connections (Felleman and Van Essen, 1991), which are thought to be crucial to the image processing computations occurring therein. Neural network architectures in which the activity in a node can feed back and influence activity in a previously active node are called recurrent. Whereas feedforward CNNs are universal function approximators, recurrent CNNs are universal dynamical system approximators, able to represent the dynamic evolution of a system, utilizing its internal network state as a form of memory that allows it to detect temporal as well as spatial patterns in its input and capture those patterns in its output (Kriegeskorte, 2015). This dynamical system-representing ability of recurrent CNNs evokes both the anatomical architecture and dynamical function of biological neural networks. Furthermore, the lateral and feedback connections of recurrent neural networks are thought to play an important role in noise reduction and occlusion compensation in object images degraded by noise or clutter (Spoerer et al., 2017).

Spoerer, McClure, and Kriegeskorte tested several CNN architectures on a variety of visual object recognition tasks to assess potential performance improvements afforded by top-down and lateral connections (Spoerer et al., 2017). The researchers compared the performance of four types of CNN architectures: exclusively feedforward, feedforward with lateral connections, feedforward with feedback connections, and feedforward with lateral and feedback connections. The two perceptual tasks the researchers pitted against the CNNs were digit discrimination tasks that occluded target digits in two different ways: with other digits and with digit fragments. These forms of occlusion were designed to mimic how occlusion of an object in a natural image results in degraded visual input for the object. In natural scenes, task-relevant objects are often partially occluded by different task-irrelevant objects that co-occur with the target object, or by textures that resemble parts of the target object because of similar visual features. For example, trying to find a corkscrew buried somewhere in a messy kitchen drawer is difficult because of the mess of random twist ties, chip clips, batteries, and rubber bands in which the target corkscrew is partially submerged. Being able to infer the presence of the corkscrew by a combination of its visible features, extrapolating its full structure behind occluding distractors, is an essential function of the human visual object recognition system. Spoerer and colleagues found that in their two occlusion tasks, the best performing model was the recurrent CNN with both feedback and lateral connections, suggesting that this recurrent architecture is essential for ecological object recognition performance.

In summary, deep CNNs represent the best current understanding of how the human visual system represents and recognizes objects. The CNN is considered the best available model both because its instantiation in computer vision applications achieves human levels of performance, and because based on our knowledge of the anatomy of the visual system and the biophysical properties of neurons, the computations performed by CNN nodes could be instantiated in biological neurons, and the algorithms embodied in the network architecture of CNNs reflects known connectivity of biological neural networks. Based on what is known about how CNNs represent and classify objects, we can infer that the human visual system builds representations of objects from initially simple to progressively more complex feature representations over successive stages of processing. The basic idea that simple features can be combined into more complex features at subsequent processing stages has been known to vision research since the early characterizations of the receptive fields of simple and complex cells, but CNNs provide a proof of concept that successively complex representations can culminate in highly specific object representations for thousands of categories, robust to widely variable physical appearances. In particular, recurrent CNNs are the best CNN models of biological visual object processing, because their network architecture closely resembles the neural networks in ventral visual cortex, and because they have been demonstrated to be robust to degraded inputs, such as occlusion, that commonly occur in natural environments (Spoerer et al., 2017). In line with that idea, recent work has revealed limitations in applicability of purely feedforward CNNs to primate object recognition systems (Rajalingham et al., 2018), so future work should evaluate the extent of the similarities between recurrent CNNs and primate visual networks in order to solidify recurrent CNNs as the most veridical mechanistic models of human object recognition.

Evidence from cognitive neuroscience

In the human brain, two major visual processing pathways show object selectivity: The ventral stream visual pathway – comprising areas V1, V2, V4, temporo-occipital (TEO), and inferotemporal (IT, or TE), and the dorsal stream, comprising V1, V2, MT, and MST (Perry and Fallah, 2014). Whereas the dorsal stream is thought to be mostly involved in the visuospatial attributes of objects that guide motor control and interaction, the ventral stream is thought to represent object form and to underlie semantic representations (Kravitz et al., 2013). Thus, ventral stream object representations are thought to integrate into cognitive control functions such as reasoning, problem solving, and planning.

That said, there is evidence that some dorsal stream areas not only possess object selectivity, but are also hierarchically organized in a manner similar to the ventral stream, with high-level dorsal stream areas responsive to object images invariant of size, position, and viewpoint, raising questions about the extent to which dorsal stream object representations are truly restricted to visuospatial operations such as grasping and tool manipulation (Konen and Kastner, 2008). Thus the function and role of the dorsal stream is a matter of ongoing concern (Mruczek et al., 2013), but the role of the ventral stream in object recognition is well-established. To that end, the object representations in the ventral stream are considered to be more elaborate than those of the dorsal stream, largely because of the recurrent neural architecture ubiquitous in ventral stream areas, strengthening the interpretation that the ventral stream is the visual system's "what" pathway (Kravitz et al., 2013; Lehky and Tanaka, 2016).

Cell electrophysiology studies in monkeys provide evidence for a functional hierarchy instantiated along the monkey ventral visual pathway. Anterior IT cortical cells, at the top layer of the hierarchy, are tuned to complex visual features, and are invariant to two-dimensional rotation, position, and size (Tanaka, 1997). Tanaka and colleagues' recording studies found that cells in anterior IT are tuned for partial-object features, rather than whole objects, so Lehky and Tanaka have proposed that object representations comprise activation of groups of complex feature-selective cells in anterior IT (Lehky and Tanaka, 2016). According to this interpretation, successively more complex features are extracted along the ventral visual hierarchy, but ultimately objects are not represented by category-tuned neurons, but rather are represented in a distributed fashion in IT cortex with a population code.

Non-human primate object recognition behavior is similar to human object recognition, as shown in a large-scale evaluation of object-level and image-level categorization behavior (Rajalingham et al., 2018), and representational similarity analysis over IT response patterns to object exemplar images (Kriegeskorte et al., 2008). Moreover, the dorsal visual pathway areas studied in monkey single-unit recording studies are thought to have homologous partners in the human visual stream (Lehky and Tanaka, 2016). The commonalities among humans and monkeys suggest that object recognition mechanisms in the ventral visual pathway are conserved within primates, and so the findings that anterior IT cells are tuned to complex partial-object features and that populations of IT cells code object identities likely apply in humans. However, the same neural recording methods cannot be applied in healthy human participants, so neuroscientific study of human object recognition mechanisms has been restricted to non-invasive methods such as fMRI and EEG.

Some of the first significant neuroimaging studies of the function of human IT cortex revealed relatively large cortical areas selectively responsive to commonly encountered images, such as faces, places, and words. Findings of large clusters of category-selective neurons complicate the picture of object representations as distributed and population-encoded. Nevertheless, the findings are striking and have had a significant impact on the field. Kanwisher, McDermott, and Chun used fMRI to localize an area of IT cortex in the fusiform gyrus that is selectively responsive to images of faces (Kanwisher et al., 1997). Kanwisher and colleagues specifically designed their study to rule out several alternative possibilities that could lead to the observation of greater face image-eliciting activity in the fusiform region: attention-related factors due to the attention-drawing nature of face images, selective low-level feature processing attributable to the difference in feature composition between face images and other images, and activation due to an animacy/inanimacy distinction between faces and other images. The researchers named the face-responsive region the fusiform face area (FFA) because of its apparent selectivity for face images.

The interpretation of Kanwisher and colleagues that the FFA is a selective face image processing module has been contended. Some researchers have suggested that the area may actually function more broadly as a subordinate-level identification system. For example, Tarr and Gauthier have posited that because faces are ubiquitous in human life and are a crucial part of communication and social behavior, they need to be efficiently represented at the level of individual faces rather than at the basic-level category. The researchers contend that this automatic subordinate level category representation may be performed in the fusiform gyrus, and most human participants show selective activation to face images in the fusiform area simply because faces are the object category that disproportionately utilizes this subordinate-level identification system (Tarr and Gauthier, 2000). According to this logic, expertise in a subject matter that requires efficient discrimination of category exemplars, such as bird watching, should also be accompanied by fusiform area activation in response to images from the category of expertise, and Tarr and Gauthier presented evidence to support this prediction. Kanwisher and colleagues have refuted the interpretation of Tarr and Gauthier (Kanwisher and Yovel, 2006).

Despite varied interpretations of the functional reason that the fusiform face area selectively responds to images of faces, neuroimaging methods similar to those that localize face-selective activation have been successfully employed in the identification of other visual category-selective areas, such as the parahippocampal place area (PPA). Epstein and Kanwisher identified the PPA, a bilateral area in temporal cortex near the hippocampus, as selectively responsive to images containing information about the layout and context of local scenes, regardless of the number or identity of specific objects contained within the scene (Epstein and Kanwisher, 1998). In further examination of the PPA, Epstein, Harris, Stanley and Kanwisher found that activation in the PPA is not associated with place recognition, movement through places, or more general memory mechanisms, but instead is highly specific to coherent scene layouts and encoding of novel place information (Epstein et al., 1999). In more recent work, the evidence for category selectivity in FFA and PPA has been supported by Mur and colleagues, who analyzed single-image fMRI responses in these areas to demonstrate category selectivity independent of category-averaged BOLD signals (Mur et al., 2012). Henriksson, Mur, and Kriegeskorte have further refined our understanding of the PPA's selectivity – Henriksson and colleagues' work suggests that PPA is more selective for scene texture and semantic meaning, whereas the nearby occipital place area (OPA) is selectively responsive for the geometry and spatial layout of scenes (Henriksson et al., 2019).

Findings from human neuroimaging studies suggest ways that human IT cortex represents basic level object categories. Faces and places may not seem like objects according to the vernacular usage of the term “object,” which usually connotes a solid, three-dimensional entity. Indeed, researchers studying the properties of human IT cortex sometimes make a distinction between “face” and “object” recognition (e.g., Kanwisher et al., 1997). But as discussed in an earlier section of this paper, we can define a visual object to be a collection of bound visual features whose representation is robust to changes in its low-level visual properties such as size, position, and feature set. Faces and places can therefore be considered objects within this framework.

IT cortex has also been implicated in object representations for categories other than faces and places. Unlike these special categories, selective cortical activation evoked by extraneous object images cannot be localized to highly specific areas for most object categories. Haxby and colleagues analyzed multi-voxel patterns elicited in IT cortex by various categories of objects including faces, cats, and human-made items like shoes, tools, and chairs, and found that voxels reliably activated by non-face objects covered a large swath of IT cortex, rather than being clustered into focal areas (Haxby et al., 2001). This finding suggests that the population code for object representations generally is instantiated in a distribution across a large cortical region, and that object category-selective areas are not all focal cortical modules.

In further support of this interpretation, Kriegeskorte and colleagues have investigated the categorical structure of IT cortex response patterns using representational similarity analysis (RSA) (Kriegeskorte et al., 2008). Kriegeskorte and colleagues compared patterns of IT activation in response to individual object images including faces, body parts, animals, and inanimate objects, and found that unlike in early visual cortex, IT response patterns showed broadly similar response patterns within object category. These results were maintained even when FFA and PPA were excluded from the analysis region of interest, implying that face and place selectivity was not driving this category formation. Further work in this vein revealed that the category boundaries evinced by the RSA analysis correlated with decision boundaries when participants were asked to make explicit category distinctions for individual object images, underscoring the behavioral relevance of the object category representations instantiated in IT cortex (Carlson et al., 2014).

Thus, from neuroimaging studies that revealed visual areas selective for faces, words, and body parts, and areas that seemed to be commonly selective for many other types of objects, with different multi-voxel patterns of activation associated with different object categories, one interpretation is that IT cortex is divided into modules for domain-specific processing of visually important and commonly encountered objects, such as faces, and modules for general processing that represented all other categories of objects with a population code of partial-object features (Grill-Spector, 2003; Grill-Spector and Weiner, 2014).

One prediction that could be made upon this conceptual foundation is that because object representations are distributed and encoded with population-level activity patterns, recognition and discrimination of two objects that share similar representational activity patterns should be more difficult than for two objects whose activity patterns do not overlap. Indeed, researchers have found that this is the case (Cohen et al., 2014). Furthermore, Cohen and colleagues have shown that the similarity of underlying IT patterns for two objects affects whether the two objects will be able to access observer awareness at all (Cohen et al., 2015).

Intriguingly, these findings suggest a possible reason that behaviorally critical objects such as faces and words are bestowed with highly selective and focal cortical areas for their representation: The developmental pressure to be able to recognize and process face information in cluttered natural scenes could form representational hardware that infrequently gets co-opted for constructing representations of other, lower-priority objects. For example, because faces are frequently behaviorally relevant in human behavior – representing a critical interpersonal communication channel, supporting theory of mind processes, displaying threat signals, and serving many other crucial functions – if a large portion of the distributed population code that represented faces was shared with the representative population for something equally commonplace but relatively irrelevant to human behavior like pigeons, then whenever a pigeon and a face were present in the same scene, the visual system might exclusively represent the pigeon, precluding the face information from being computed for important behavioral cues and even from reaching conscious awareness. Having a population of partial-object feature detectors that are solely dedicated to the partial-object features used to construct face representations would prevent this maladaptive outcome from occurring. And it can be hypothesized that these dedicated feature-encoding neurons cluster together into large, focal cortical areas, such as seen in the FFA, because such clustering is the optimal metabolic arrangement when their activity and network properties are relatively isolated from the rest of object-selective cortex.

The notion of a population code for object representation has thus been firmly established, prompting researchers to investigate its spatial dimensions in occipitotemporal cortex. In other words, researchers have sought to understand how the large swaths of object-responsive cortex are organized.

Konkle and Oliva found evidence that object representations in occipitotemporal cortex are organized according to their real-world size (Konkle and Oliva, 2012). The researchers showed that while both small and large objects activated voxels throughout the functionally defined region of interest in occipitotemporal cortex, small objects preferentially activated more lateral sites in occipitotemporal cortex while large objects activated more medial sites. These results occurred irrespective of the retinal size of the object image and even whether the object was physically presented or only present in mental imagery, suggesting that this differential size representation reflects object concepts and not feedforward visual information. In further work along these lines, Konkle and Oliva found that real-world size is not differentially represented in occipitotemporal cortex for animals, suggesting an orthogonal dimension of animacy (Konkle and Caramazza, 2013). Konkle and Oliva concluded that real-world object size and animacy reflect fundamental dimensions along which object representations are encoded, although the reason that such arrangements of object representations are instantiated in the brain is an open question. For example, the differential preference for real-world size may not reflect a fundamental importance of size itself, but rather geometrical properties or visual features that covary with objects of different sizes. In support of this idea, Long, Yu and Konkle have shown that the real-world size and animacy preferences in object-selective cortex could be replicated even when object identities were obscured but mid-level visual features belonging to those objects were preserved (Long et al., 2018).

In summary, evidence from cognitive neuroscience depicts human object recognition as occurring mainly in the ventral visual pathway, and as instantiated with a population code covering large cortical areas rather than being built in a purely feedforward fashion along successive steps in a ventral caudal-to-rostral hierarchy that culminates in a singular object-tuned abstraction (Haxby et al., 2001; Konkle and Caramazza, 2013). Recurrent neural architecture, comprising feedforward, feedback, and lateral connections between ventral visual areas, are essential to the formation of object representations in the ventral visual pathway, and undermine the simpler, early model of ventral stream object representation that describes a chain of visual processing in which successively more complex features are extracted from visual input at each stage (Kravitz et al., 2013). Nevertheless, category representations have been shown to be hierarchically constructed (Carlson et al., 2013; Cichy et al., 2014), suggesting that mechanisms of object representation incorporate both distributed encoding and hierarchical processing. Thus, evidence from cognitive neuroscience supports the recurrent convolutional neural network model of object recognition (Spoerer et al., 2017).

Object-Based Attention

Attention is predominantly studied in a visuospatial context. The beginnings of the cognitive neuroscience of attention operationalized attention in terms of a cued spatial attention design (Posner, 1980; Posner et al., 1980), and since then, cued spatial attention has been a dominant experimental paradigm. But implications of the construct of attention, theoretical concerns about how attention may be implemented, and neuroscientific evidence of attentional mechanisms all suggest that attention should be able to operate over objects, or even primarily act on object representations.

The construct of attention is based on the functions that attention is thought to serve for organisms enacting adaptive behaviors in natural environments. One of these functions is facilitating the identification of behaviorally relevant objects in the environment. For instance, in a rainforest abundant with shimmering dewy fronds, swaying vines, and light beams spilling through overgrown canopies, being drawn towards any momentarily salient stimulus would be an ineffective way to find a delicious ripe avocado. A top-down system that selects avocado-related stimulus information for preferential processing would be critical for survival in this kind of world. Moreover, because any given object could display an infinite variety of low-level visual features, such as its position in space, its size, and its appearance under variable and dynamic lighting conditions, it would theoretically be most efficient for a natural object-search system to be able to act on high-level object representations that are invariant to the low-level visual features, rather than having to activate all possible low-level visual feature representations that might be associated with the target object in the present scene. This latter scenario would be especially maladaptive, because it would lead to the selective enhancement of non-target objects that inhabit a similar area of low-level visual feature space as the target object, for all possible appearances of the target object. The resulting effect would hardly resemble the kind of sensory filtering associated with attention.

As discussed previously, top-down attention control signals are thought to influence the processing of visual areas that receive input from different information sources, whether those information sources are spatial locations, sizes, specific or visual features. If all possible control signals were transmitted to their respective sites of action over dedicated anatomical channels, the number of necessary anatomical connections would be exceedingly large. Therefore, this kind of architecture seems unlikely to be implemented in the brain (Buschman and Kastner, 2016). To severely constrain the number of necessary anatomical links, attention control signals may primarily target object representations, because object representations contain information about the object’s lower-level visual features. Thus object-based attention may be fundamental mechanism of action for the top-down attention control system, and object representations may be the primary units over which attention operates.

Findings from neuroscientific studies of attention’s mechanisms also suggest that attention may be able to directly target high-level object representations. For example, the finding that the pulvinar may be critically involved in the attentional regulation of communication between cortical areas (Saalmann et al., 2012) could be applied to an object-selective attention mechanism, because the pulvinar is highly connected with the cortex – directly connected cortical areas are also indirectly connected via the pulvinar. Thus, information transmission along specific circuits in the ventral visual pathway that subserve targeted object representations could in theory be selectively enhanced through the pulvinar.

I will now turn to the topic of object-based attention in greater detail. In this section, I will describe findings from cognitive neuroscience that have begun to elucidate the neural mechanisms of object-based attention, and propose a synthesis of object recognition research and object-based attention research. I aim to answer the question of how object representations serve as substrates for attention, based on the current understandings of how attention operates at a neural level and of how the visual system represents objects. Then I will propose a novel hypothesis that the factors thought to guide attention via top-down control, such as goals and task, may instead be conceptualized as an extension of object recognition mechanisms into increasingly more abstract realms. This hypothesis synthesizes object-based attention and top-down attention control into a singular phenomenon. I call this hypothesis the Natural Object Attention Hypothesis. But first, I will provide a more detailed description of what is meant by the term “object-based attention.”

What is object-based attention?

The concept of object-based attention builds on the concept and operational definition of attention more broadly construed. Attention is defined as a selective enhancement of task-relevant information sources to the exclusion of other, potentially targetable information sources (Luck and Vecera, 2002). In the same way that visual, auditory, somatosensory, spatial, feature-based, and other narrow forms of attention can be understood as refinements of this broad definition, object-based attention can also be accommodated: Object-based attention is a selective enhancement of task-relevant object information to the exclusion of other, potentially targetable objects. What differentiates object-based attention from other forms of visual attention is that rather than selectively enhancing information from the region of space containing the target object or the specific visual features constituting the object, object-based attention specifically acts over high-level object-representations themselves. Thus, for example, there is a conceptual distinction between the act of covertly attending to a region of space and attending to the object that is situated in that region of space, even if the qualitative experience is difficult to distinguish between these two scenarios – the heightened awareness of the object in that region of space and its visual appearance may be experienced identically even if the control mechanisms that guided attention to produce those enhanced perceptions were different. Thus the conceptual distinction is primarily made on the basis of the putative top-down control and selective mechanisms operating in the brain. In comparison with spatial attention, object-based attention invokes a targeted enhancement of object representations directly, instead of an indiscriminate enhancement of all visual input coming from a region of the visual field.

Object-based attention can be operationalized and studied within the same framework as attention more broadly construed. For example, a cuing paradigm has been used to study object-based attention (Noah et al., 2020). In this study, predictive cues indicated the upcoming appearance of an object image from one of three categories, and participants were instructed to discriminate the subordinate-level category of the object image – for example, if it was a face image, participants would have to indicate whether it was a male or female face. On a fraction of invalid trials, the target object image would belong to an uncued category, and by analogy with the cued covert spatial attention paradigm, an attentional reorienting would occur from the cued object category to the presented object category, so that the behavioral task could be performed. Just as in cued spatial attention experiments, this reorienting would have a reaction time or accuracy cost that could be measured, providing empirical access to object-based attention. In a similar manner, search, dual-task, and filtering paradigms (Luck and Vecera, 2002) could all be constructed for object-based attention tasks.

Cognitive neuroscience of object-based attention

From the definition of object-based attention provided above, a prediction that immediately follows is that attention directed to an object spreads readily over that object, enhancing all its perceptual attributes. Behaviorally, this means that if two perceptually distinct objects are concurrently presented, and a location on one of the objects is cued, then response times and discrimination accuracies should be faster when a subsequent target appears on the object that contains the cued location compared to when the target appears outside the cued object. This prediction has been borne out in several influential early studies of object-based attention, and has been strengthened by important recent work.

Egly, Driver, and Rafal showed that target detection is faster when the target appeared on the cued object than when it appeared on an uncued object or elsewhere, even when the distances between the cue and the targets in the two conditions were held constant (Egly et al., 1994). Moore, Yantis, and Vaughan extended this finding by showing that the “same-object advantage” was also active even when the two objects were partially occluded by a third object, such that the cue and same-object target were separated by the occluding object in the same-object condition (Moore et al., 1998). Thus Moore and colleagues inferred that the mechanism that enabled attention to spread within perceptually defined units was not entirely driven by low-level properties like line continuity and contour detection, but rather occurred over a more abstract object representation that formed after early visual features were parsed.

These studies from Egly and colleagues and Moore and colleagues were conducted with impoverished stimuli. In other words, the objects that the researchers used to demonstrate their same-object advantage were not very natural looking. The objects presented were plain rectangles on blank backgrounds. This type of impoverished visual stimulus is commonplace in cognitive neuroscience and psychology research for good reason – stripping psychological phenomena to their minimally necessary and sufficient components provides experimental control, a basis for sound logical inference, and a descriptive understanding of the phenomenon that can potentially be modeled mathematically. But in the case of object-based attention, a construct that is partially motivated by its applicability in real-world scenarios, plain line figures isolated against blank backgrounds hardly seem to capture the essential character of the natural phenomena under investigation. Hence the initial findings from Egly and colleagues and Moore and colleagues that attention spreads throughout perceptually defined objects could arguably not be considered ecological. However, recent work has extended the minimal form of the same-object advantage effect to veridical visual stimuli. Malcolm and Shomstein have shown the same pattern of same-object advantage effects from images of natural scenes containing multiple discrete objects (Malcolm and Shomstein, 2015). This study links early theories that attention spreads within perceptually-defined discrete units to a more categorical and semantic context.

Critically, in the studies of Egly and colleagues, Moore and colleagues, and Malcolm and Shomstein, participants were not explicitly instructed that the appearance of the cue indicates that the entire object containing the cue should be attended. Rather, the cue was designed to capture attention exogenously and not be instructional, thus providing a logical basis for the claim that attention spreads automatically throughout the visible perceptual features of the cue-exhibiting object. Alternatively, rather than spreading throughout the visible boundaries of the cue-exhibiting object like water filling a container, attention may be thought to *fundamentally operate over* the cue-exhibiting object, and objects in general. In support of this hypothesis, O'Craven, Downing, and Kanwisher used fMRI to show that attention to a behaviorally relevant visual feature, such as movement, concomitantly increased activity in both the visual area selective for the relevant feature and also the area selective for the category of the object that exhibited that feature in the experimental stimulus (O'Craven et al., 1999). In their experiment, participants were presented with stimuli that were composed of two overlaid semi-transparent object images – a face and a house – and instructed to attend to motion. Only one of the two object images in the overlay would show the motion. BOLD signal increases were detected in the motion-responsive area MT as predicted by previous work on attention's neural effects, but BOLD signal increases were also seen in either PPA or FFA, depending on which image moved. This finding is incommensurable with models of attention requiring that attention uniformly enhances processing of all information at an attended location, because the stimulus images were overlaid and thus spatially coextensive. If attention enhanced all the visual information present at the attended location, BOLD signals in PPA and FFA would have changed to the same degree. Instead, attention to the low-level visual feature of motion seemed to be inextricable from the object displaying that visual feature. Thus the finding of O'Craven and colleagues is supportive of an object-based model of attention, according to which all attributes of an object are selected by attention simultaneously.

ERP work from Woodman, Arita, and Luck supports the idea that attention fundamentally operates over objects (Woodman et al., 2009). In their study, the researchers showed that the N2pc ERP component, thought to index lateralized shifts of attention, was only observable when shifts of attention were directed to regions of space that contained placeholder objects, rather than being empty and structureless. The researchers interpreted this to mean that anticipatory shifts of attention to cued locations could occur when objects in the cued location were present to serve as anchors, and that when this was the case, the behavioral effects of attention could indeed be attributed to neural processing occurring in advance of target presentation, rather than only upon target presentation. Woodman and colleagues' study utilized what might be considered impoverished objects, so future work to replicate their findings with natural objects would be beneficial for an account of the primacy of object-based attention.

Indeed, recent work has emphasized that understanding the way attention operates in the real world must center on object-based mechanisms and natural object images. For example, humans can rapidly detect objects in natural scenes, suggesting that the visual system possesses a selection mechanism for objects that can operate independently of spatial attention. This rapid object detection behavior contrasts with predictions that may be made on the basis of foundational visual search research involving impoverished stimuli. Early attention research showed that in a visual search array of simple objects such as letter Ts, participants can immediately find a target T if it is a different color than the other Ts in the array. The item displaying the singleton feature pops out from the entire scene (Treisman and Gelade, 1980). This is taken to mean that one-feature search can be conducted in parallel over the entire search array. However, as soon as the conjunction of two features, such as shape and color, are necessary for identification of the target, search times scale linearly with the number of distractors in the array. For instance, detection of a green letter T among green and red Ts and Ls cannot be achieved instantaneously, but rather the amount of time taken to find the target feature conjunction scales linearly with the number of search items, suggesting that in order to perform this search, each item must be inspected individually, requiring attention to move serially across the visual display. From this foundational research, one might predict that in a search task involving the detection of a natural object in a natural scene, such as a car on a typical street scene, search would be extremely slow, because the target is a conjunction of a multitude of feature dimensions, and each of those feature dimensions have to encompass more specific features than will be displayed by the target object. Searching for a car in a crowded naturalistic scene will not involve searching for a particular assembly of oriented lines, colors, textures, sizes, and motions, but many such possible combinations that can be produced within these dimensional constraints: a red hatchback, a blue minivan, a white convertible, etc., driving, stationary, on blocks, etc. Counterintuitively in the light of these prerequisites, but somewhat obviously given that such forms of visual search are common occurrences in everyday life, participants engaging in such tasks can detect the natural image targets as if they were conducting parallel search across the whole scene, generally detecting their targets within 100 msec after image onset (Peelen et al., 2009; Peelen and Kastner, 2011, 2014).

Peelen, Li, and Kastner used fMRI to investigate the neural processes underlying this rapid object detection behavior that seems to contrast with foundational research on simple stimuli (Peelen et al., 2009). In their experiment, the researchers instructed participants to search for either people or cars in a large set of natural outdoor scene images. The researchers analyzed multi-voxel patterns, from functionally localized object-selective cortex, evoked by scene images. These patterns were compared against similar patterns obtained from a separate task in which isolated objects were presented at the center of fixation. In comparing object cortex patterns between the natural scene search task and the passive object viewing task, the researchers found greater correlations for within-category multivoxel patterns than between-category patterns, suggesting that the body and car information was only extracted from natural scenes and represented in high-level object-selective cortex when these objects were being actively searched for. Furthermore, the researchers were able to show that these high-level object representations were formed even when the target objects were only present in scene images outside spatially cued regions. The researchers interpreted their results as reflecting an attentional pre-activation of object category representations that biases the processing of subsequent visual scene input.

In further related work, Peelen and Kastner found that when participants anticipated a natural scene containing an object from an attended category, the multi-voxel patterns elicited in object-selective cortex closely resembled the patterns evoked by the actual presentation of object images from that category (Peelen and Kastner, 2011). This finding provides evidence that attentively anticipating an upcoming object image, as in the context of a visual search, increases baseline activity in the population of cells that encode and represent that object category. This inference extends work on the mechanisms of visuospatial attention to the domain of objects. Furthermore, Peelen and Kastner observed that multi-voxel patterns in medial prefrontal cortex also were distinguishable and correlated with attended object category, suggesting the involvement of this area in the selective enhancement of object representations in occipitotemporal cortex.

In addition to the work of Peelen and Kastner (Peelen and Kastner, 2011), other studies investigating the sources of attentional control over objects have revealed the involvement of frontal areas similar to those involved in spatial and feature-based attention. For instance, Serences and colleagues used fMRI to look for the sources of top-down control over object-based attention in an experimental design that required participants to switch attention between a face and a house in a series of overlaid face-house images (Serences et al., 2004). The researchers found that activity in FFA and PPA was modulated according to the attended object category, in line with previous findings showing heightened neural activity in areas selective for attended objects (O'Craven et al., 1999). Furthermore, the researchers observed that switches in attention between object categories corresponded to transiently heightened activity in parts of the dorsal frontal cortex and posterior parietal cortex that have also been implicated as sources of top-down attentional signals in other domains (Corbetta and Shulman, 2002). Thus the researchers concluded that the dorsal attention control network was involved in the effortful switches between attending to different object categories, just as it has shown to be involved in attention to low level features.

Baldauf and Desimone used MEG and fMRI to show that the inferior frontal junction (IFJ) was involved in the top-down maintenance of attention to one of two categories of objects (Baldauf and Desimone, 2014). In Baldauf and Desimone's experiment, participants were presented with a continuous stream of face and house images, with each object category image presented at one of two flicker frequencies. The two object categories' flickering image streams were interleaved and out of phase, so the two object categories could be presented at the same spatial location. Participants were instructed to attend to one of the two object categories. MEG data collected while participants performed this task could thus localize frequency-tagged activity belonging to each object category, and as expected, from the MEG data face-related frequency activity could be localized to the FFA and house-related activity to the PPA. In IFJ, the researchers observed MEG signals at the frequency of the attended object category's presentation rate, indicating that the IFJ was functionally linked to the activity in the object-selective areas and supported object-based attention in this task. Baldauf and Desimone concluded that the frontal control areas involved in object-based attention were similar to those involved in spatial attention, but whereas the FEF issues spatial biasing signals, the IFJ issues object biasing signals.

In summary, evidence of a behavioral same-object advantage in visual information processing (Egly et al., 1994; Moore et al., 1998; Malcolm and Shomstein, 2015), and neuroimaging evidence that attention enhances all attributes of an attended object even when those attributes are not task-relevant (O'Craven et al., 1999) supports the idea that the primary units of attention are objects. Object-based attention mechanisms may be fundamental to how attention operates in natural environments (Peelen and Kastner, 2014). Neuroimaging studies have revealed that the same networks involved in top-down control of spatial and feature-based attention are also involved in attention to objects (Serences et al., 2004; Baldauf and Desimone, 2014), raising new questions about whether attention to regions of space or low level features are directly attributable to activity in the dorsal attention network, or whether instead attentional modulations of low-level features occur via their interaction with object representations.

How attention operates on object representations

The recurrent neural architecture of the ventral visual pathway (Kravitz et al., 2013) supports the implementation of object-based attention. And from this implementation can be extrapolated the properties and behaviors observed by the psychological and neuroscientific studies heretofore described.

Object-based attention processes begin in the dorsal attention network (Serences et al., 2004; Baldauf and Desimone, 2014). This network translates task instructions, or behavioral goals, into signals that descend the processing hierarchy into object-selective cortex, and activate the distributed population of neurons that jointly encode the targeted object category. Crucially, because of the recurrent architecture of this visual processing system, it may be hypothesized that attention control signals do not need to transmit simultaneously to all nodes in the target object-representative network. Rather, activation of any node or subset of the target network will spread to the connected nodes. This property of the object representation system, supported by recurrent connections, potentially answers the question of how the anatomical limitations of the dorsal attention network could possibly allow connection to all targetable sensory sites. The space of objects that can be selectively attended is certainly enormous, but when multiplied by the number of spatial positions, sizes, and feature compositions that could be combined with any object, the number of precise anatomical connections that would be required to support a one-to-one mapping would be prohibitively large for cortical instantiation. Rather, by limiting the structural connections between the dorsal attention network and visual cortex to just the number of objects that can be selectively attended, and then relying on the recurrent architecture of the ventral visual pathway to bias activation of low-level visual features that constitute the attended object, all forms of attention can realistically be accommodated within limited anatomical real estate.

Buschman and Kastner have proposed a theory of attention that aligns with this object-based picture. According to their theory, higher-order attention control areas issue broad, non-specific signals to sensory sites, and the local circuitry of those sites works in concert with the top-down signals to enhance processing of selected sensory information (Buschman and Kastner, 2016). Buschman and Kastner propose that broad top-down signals lead to enhancements of precise object representations by their stimulation of local sensory area circuits that have a pattern completion functionality. Pattern completion ability is thought to emerge from the ability of the visual system to encode natural image statistics from incoming visual input (Simoncelli and Olshausen, 2001).

A model of object-based attention in which object representations are selectively activated, combined with the current understanding of how objects are represented in the ventral visual pathway as distributed, population-encoded activation patterns, leads to some predictions about neural data that should be obtainable under different experimental conditions. For instance, because the population code for object representations is distributed over large swaths of visual cortex, including low level visual occipital areas, top-down attention to a specific object category might be expected to activate not just the multi-voxel patterns in object-selective cortex but also in early visual cortex that extracts the low-level features associated with that object. Thus differences in activation patterns in these low-level areas between object attention conditions should be apparent when the object categories being contrasted possess dissimilar features. Faces and houses are composed of highly dissimilar low-level visual features: Faces contain predominantly rounded shapes whereas houses exhibit strong linear features with many vertical and horizontal edges. Thus comparing object-based attention between these two object categories should yield observable multi-voxel pattern differences in early visual cortex. Cohen and Tong tested this prediction, examining the effects of object-based attention on lower-level visual areas V1 – V4, and found that indeed, multi-voxel activity patterns in V1 – V4 were discernable between two object attention conditions involving faces and houses (Cohen and Tong, 2015). Future work should test the prediction that the degree of decodability in V1 – V4 correlates with the dissimilarity in low-level visual feature composition for the classes of objects being attended.

An account of attention control signals that lead to the selective enhancement of processing of object-specific visual input, by a mechanism such as an increase in baseline firing rates of a population of IT neurons that encode the targeted object (Peelen and Kastner, 2011), squares with scenarios in which object-based attention is deployed in anticipation of an object that is not yet in view. However, can this model account for cases of object-based attention when the objects in question are already visible, but attention spreads over the object after an event such as an exogenous cue? Early studies of object-based attention emphasized that findings such as the ones of Egly, Driver, and Rafal (Egly et al., 1994) suggest a visual processing stream in which the continuous, undifferentiated field of visual input is parcellated preattentively into discrete objects, on the basis of Gestalt grouping principles (Behrmann et al., 1998). But does this interpretation hold up in light of a more contemporary understanding of object representations in the brain, and attention's neural mechanisms? This question is open to future investigation, but one explanation is plausible within the framework described here. In a scenario such as the one designed by Malcolm and Shomstein (Malcolm and Shomstein, 2015), a natural visual scene containing multiple discrete objects, such as a grassy lawn with chairs, is presented to an observer. Without any instruction to attend to a particular object, attention may be drawn to the most salient or meaningful object visible in the scene. But when an exogenous cue, such as a transient luminance increase, occurs on one of the chairs, one of the low-level features of that chair, such as a particular spatial location, is suddenly made more salient. This is reflected physiologically by a transient increase in neural activity in a visual feature detector that is bound to other feature detectors, both low-level and high-level, partial-object feature detectors. Thus the transient activation spreads throughout the distributed network that jointly represents the chair in that particular image, leading to an enhanced activation of the entire chair representation. This enhanced activation is observed as exogenously cued attention to the chair, and supports subsequent target presentation when it occurs on the same chair, but not when it occurs on a different object, even if the location of the target on that object is equidistant from the cue location, because that object's distributed representation has not been exogenously activated by the cue.

The Natural Object Attention Hypothesis

An insurmountable theoretical problem in the cognitive neuroscience of attention is the infinite regress invoked by the notion of voluntary, top-down attention control. Classically, top-down attention is subdivided into two major conceptual components. One component is the effect of attention on the sensory processing sites. The second component is the origination and transmission of attention control signals from top-down sources. As the present chapter described earlier, this site-source distinction has been the basis of fruitful research into attention's neural mechanisms, providing a framework for understanding the processes that unfold in the brain when behavioral goals or task instructions require specific enhancements of sensory input. However, an intractable conundrum follows from this classical model. Namely, if voluntary attention is attributed to a putative source of attention control signals, such as the dorsal attention network, what is the proximal cause of activity in the source site? Activity in the source site of attention control signals would have to be localized to a new area – the source of attention-control-signal control signals. Attribution of voluntary exertion of attention would then shift to this newly identified second-order source, but the same problem would recur: What third-order source transmits signals to the second-order source? Hypothetically, a researcher could follow this logical train on an infinite path. Thus the source-site distinction ultimately is a barrier towards understanding attention in the brain.

The Natural Object Attention Hypothesis is a novel conceptual framework that sheds the source-site distinction. In place of the classical dichotomous framework, top-down attention is reconceptualized as an aspect of how behavioral goals are fundamentally extensions of sensory processing. The Natural Object Attention Hypothesis might solve several longstanding yet understated puzzles in cognitive neuroscience, including the infinite regress of voluntary attention control sources, the problem of object ontology, and the Resource Question in attention research. Because the present paper has reviewed work from visual attention research, I will focus the following discussion of the Natural Object Attention Hypothesis on the domain of vision, but the hypothesis does not make any strong claims that exclude sensory processing from other domains.

Object recognition research has revealed that objects are represented in the ventral visual pathway as activation patterns over a distributed population of neurons. This population code is built up over successive stages of visual processing, in which different visual information at different levels of abstraction are extracted. Crucially, there is no “final layer” at which a hierarchical process of abstraction and feature extraction culminates in a singular object representation. Rather, object representations are embedded in the network connections of a distributed population of neurons, and thus the population code relies on resources that belong to a shared pool, as demonstrated by research showing that successful representation of multiple concurrently presented objects depends on the dissimilarity of the underlying neural patterns supporting those objects (Cohen et al., 2014, 2015). Moreover, evidence is accumulating from the cognitive neuroscience of attention that suggests that attention primarily operates over object representations (O'Craven et al., 1999; Peelen and Kastner, 2014; Buschman and Kastner, 2016).

This object recognition system can thus be thought of as just a part of a larger system whose elemental function is to compute abstract information from sensory input. This system could support task set and behavioral goals in the same way that it represents visual objects. If we consider that after objects have been successfully recognized in the ventral visual pathway, there are yet further layers in the recurrent convolutional neural network that continue to perform the same basic computations, and that these layers extend throughout the brain, including into dorsal frontal and posterior parietal cortex, we may see how activation patterns in the dorsal attention network could actually be population-encoded representations of abstract properties of visual objects, such as their relevance to the task at hand.

But how does reconsidering the formerly designated "source" of attentional control signals as part and parcel of visual object recognition systems account for the empirical evidence that it transmits signals to sensory processing "sites" that modulate activity therein? According to the model of the brain's object recognition system as a recurrent CNN, activation and computation occurring at any layer has the ability to influence both upstream and downstream computations. Although research into the viability of recurrent CNNs as models of biological object recognition is still in early stages, recurrent architecture is thought to be necessary for human-like object detection performance (Spoerer et al., 2017; Rajalingham et al., 2018). Thus the architecture of a deeply interconnected CNN, in which high-level layers freely influence lower-level layers, could capture the network properties and the time-varying directional influences of the dorsal-attention-network-visual-system complex.

This new conceptual framework for attention motivates broadening the concept of an object. As described earlier in the present chapter, objects in vision science are generally considered to be discrete, bound collections of visual features, none of which are individually necessary or sufficient for objecthood; as such, objecthood is a property that is invariant over a large space of feature combinations. Defining an object in this way is an attempt to formalize the folk-psychological and vernacular term “object.” But there is no strong reason to believe that vernacular terms map onto neural processes, and actually, in this case trying to do so spawns a philosophical puzzle about object ontology. David Marr grappled with this puzzle in his book *Vision* (Marr, 1982):

> *“Is a nose an object? Is a head one? Is it still one if it is attached to a body? What about a man on horseback? These questions show that the difficulties in trying to formulate what should be recovered as a region from an image are so great as to amount almost to philosophical problems. There is really no answer to them – all these things can be an object if you want to think of them that way, or they can be part of a larger object.”*

Because language and meaning are emergent properties of behavioral interactions within large human populations, it is extremely unlikely that any particular word isomorphically links to a real phenomenon occurring inside an individual person's brain. In light of this pessimism about the utility of vernacular terms for neuroscience, we can reconfigure the classically defined object into what I propose to call a Natural Object. A Natural Object is defined by all the same qualities as an object, but includes not just visual features but also more abstract qualities such as affordance and momentary behavioral relevance. Crucially, a Natural Object does not refer to anything like a Platonic "ideal" object. In fact, the Natural Object conjecture as a philosophical stance precludes the possibility that ideal objects exist in any meaningful way, and thus, a brain's internal object representation is not a reflection of an objective, external state of affairs. Instead, Natural Objects only exist by virtue of their momentary behavioral relevance and their relationship to the observer.

The Natural Object concept and the Natural Object Attention Hypothesis is motivated by an emerging understanding of the fundamental function of the visual system as an approximator of natural image statistics. New lines of thinking propose that the essential function of the visual system is to extract the statistics of natural images (Simoncelli and Olshausen, 2001). The term "natural images" often refers to static images, cross sections of the unremitting stream of visual input from the environments in which our brains develop, but an important component of the actual natural visual input that our brain receives is its dynamic and continuous character, especially as we move through the world and interact with our environment. Statistics from this dynamic visual input are efficiently captured in the visual system as object representations. Under the Natural Object Attention Hypothesis, behavioral relevance and momentary task goals are also considered to emerge from environmental statistics, and so the binding of task into visual features to form Natural Object representations reflects the fundamental unsupervised learning procedure implemented by the brain.

Thus the infinite regress problem implicit in the source-site model and an object ontology question arising from the mapping of the brain's object representations onto external states of affairs can both be addressed within the Natural Object Attention Hypothesis. Another difficult issue related to attention research that is tractable within the framework of the Natural Object Attention Hypothesis is what I refer to as the Resource Question. In short, attention is often construed as capacity-limited: The sheer amount of incoming sensory input is thought to exceed the processing bandwidth and computational resource limits of our sensory systems. Such a line of thinking implies that evolution has yielded the brain – an extremely complex, highly-effective, goal-forming, strategy-executing, abstract thought-generating, survival-oriented device – and yet has implemented attention as a solution to an accidental evolutionary byproduct: that the body's sense organs send more afferent signals to the brain than it can handle. The dubiousness of this scenario is compounded by its appeal to an unspecified computational resource that must be limited such that the sensory information allowed into the brain exceeds its capacity. No current models of attention or sensory processing more generally have put forth a plausible account of what this limited resource precisely is. This Resource Question is understated in the field of attention research, but problematic because it undercuts an often-cited conceptual motivation for attention as a real phenomenon worthy of scientific inquiry. If the hypothetical limited resource is fictional, is attention, as the putative function that exists solely to manage that resource, also a fiction? A related puzzle that likewise falls under the label of the Resource Question concerns the singular nature of attention. Namely, why is attention only able to operate over one thing at a time? Why is the brain unable to devote the same level of enhanced processing afforded by attention to multiple signals or information sources simultaneously, or to all of them? If attention improves the efficacy of information processing in the brain, it seems that it would behoove an organism for attention to operate over all neural processing simultaneously. The conventional framework presumes that attention is singular and not omnipotent because of the limited computational resources that it manages. However, this explanation is unsatisfying because it again invokes the unspecified resource.

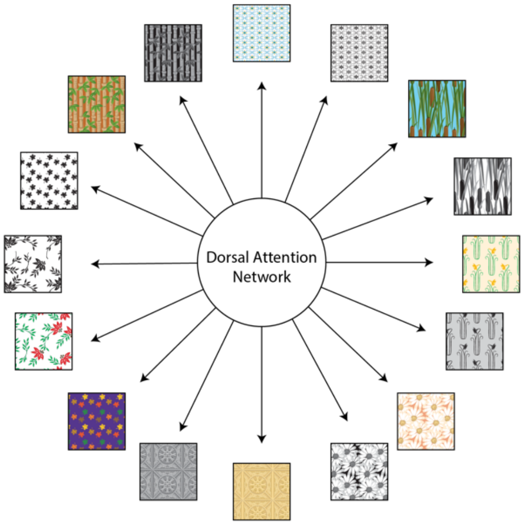
A more parsimonious explanation for the singular nature of attention that does not appeal to an unspecified resource comes from the Natural Object Attention Hypothesis. According to the Natural Object Attention Hypothesis, task goals and the behavioral relevance are bound together with visual features as Natural Object representations. The task goals represented in frontal and parietal areas such as the dorsal attention network are part of the same representational system as the object-selective areas in occipitotemporal cortex. And just as the simultaneous activation and representation of multiple objects has been shown to depend on the dissimilarity of those objects' population encodings (Cohen et al., 2014, 2015), the representation of multiple task goals and thus top-down attentional modulation may also be vulnerable to overlapping task-encoding populations. Thus phenomenologically, attention may be experienced as a singular point because the simultaneous representation of two mutually exclusive tasks in a population code, such as attention to two restricted spatial locations on opposite sides of the visual field, or attention to a face and attention to a house, is impossible.

The Specificity of Control (SpOC) Model of Attention

The Natural Object Attention Hypothesis speculates that attention is not implemented in the brain by a dichotomous site-source mode of operation. However, The Natural Object Attention Hypothesis and its corollaries are not testable with currently available neuroscience methods. Artificial recurrent convolutional neural networks are at the forefront of computational neuroscience, but are in their infancy, and not yet widely available for cognitive neuroscience researchers to access and test hypotheses about the functional behavior of such networks. Therefore, although in the future a novel framework such as the Natural Object Attention Hypothesis might prove fruitful, the site-source distinction remains useful as a method for conducting empirical research and framing the essential mysteries of how attention operates in the brain. These mysteries include questions such as: How does selectivity arise, how is selective control oriented, and what does it mean on a neural level for selected sensory representations to be enhanced or suppressed? In parsing of the neural activity accompanying attentional behavior, it is useful to think in terms of sites of attentional enhancement and sources of attentional control because of the prevailing metaphor that attention is a control system that operates over parallel streams of information processing in the sensory brain – a metaphor that is reinforced by how attention is operationalized in a laboratory setting.

The site-source distinction in conventional thinking about attention can be formalized in a framework called the Specificity of Control (SpOC; pronounced like “spoke”) model. The SpOC model seeks to explain how the pattern of activity in sources of attentional control such as the DAN represents different to-be-attended stimulus attributes (such as spatial locations, low-level visual features, and whole objects), how this information about to-be-attended stimulus attributes is transformed into or conveyed as top-down control signals, and how the functional and anatomical connectivity – both within the DAN and between the DAN and sensory cortex – enables the selection of targeted stimulus attributes or information streams.

According to the SpOC model, the DAN (the site at which attentional control signals originate) sends its modulatory signals to lower-level sensory areas via individual channels specific to the target of the modulation. This hypothetical organization contrasts with a different potential organization in which all modulatory signals are sent over a single channel – such a contrasting framework can be described as unimodal. The SpOC model postulates that the channels that carry attentional control signals are like individual spokes of a wheel radiating outward from a central hub. Each individual channel – each spoke – originates in a unique location on the hub and terminates on a specific site at the periphery of the wheel. This picture stands in for the idea that the dorsal attention network, and potentially other control areas that it complexes with, constitute a fine structure of distinguishable subnetworks – anatomical or functional – whose individual activation is the physical manifestation of an instruction to enhance a specific neural representation or sensory area (Greenberg et al., 2010; Liu and Hou, 2013). Activation of a subnetwork within the DAN is equivalent to a selective projection from the DAN to defined neural populations in visual cortex that code locations, features, and objects. If the dorsal attention network was instead domain-general (Wojciulik and Kanwisher, 1999; Fedorenko et al., 2013; Spagna et al., 2015) and/or supramodal (Shomstein and Yantis, 2004; Green et al., 2011; Wang et al., 2016; Betti et al., 2018; Salmela et al., 2018), and its activity could not be subdivided into distinct patterns that map onto specific attentional demands, then in the model illustration all the radiating lines would converge to a single point, instead of terminating on unique locations along the outside of the central hub.



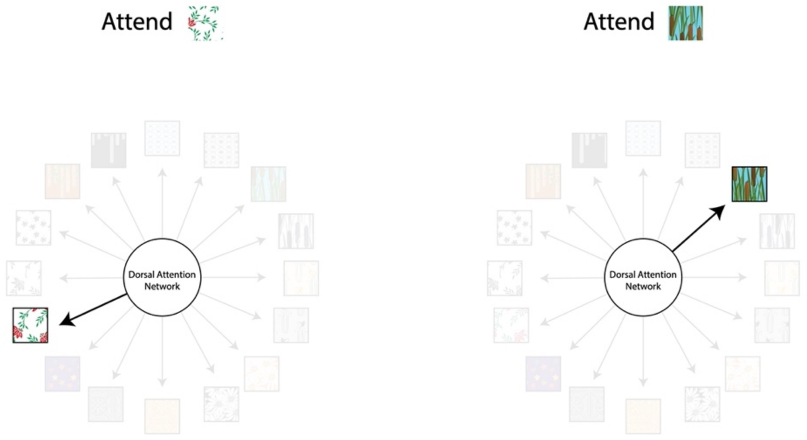


Figure 1. Cartoon illustration of SpOC model of selective attention, in which specific channels connect control network to sensory areas, resembling the spokes of a wheel. Control signals are sent selectively over these channels in order to modulate individual sensory features or representations.

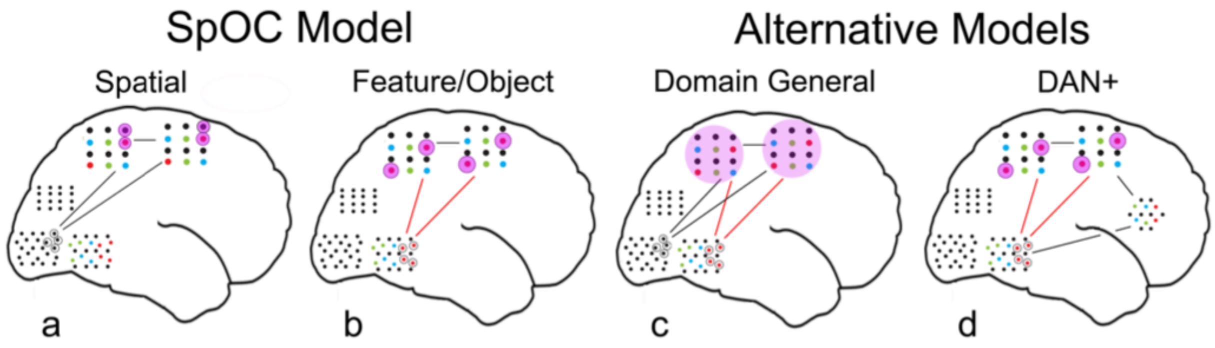


Figure 2. Reproduced with permission from G.R. Mangun. SpOC Model of attentional control as implemented in the brain, compared with alternative site-source models of attentional control. A fine structure of subnetworks in the DAN guide modulatory control of sensory areas. **(a, b)** SpOC model posits that within the DAN, different fine structures of control (translucent fuchsia circles) support different forms of attention (spatial, feature, object), and that this fine structure makes specific connections (black or red lines) with regions of visual cortex as a function of the to-be-attended visual target. **(c)** Univariate imaging has tended to support an alternative model in which the DAN represents a domain-general and supramodal attentional control system. **(d)** Recent findings in the literature raise the possibility that the DAN works together with regions outside the classical DAN (DAN+), such as the inferior frontal junction, to enable feature and object attention.

Object-based attention is an important component of the SpOC model, and moreover, the SpOC model makes several predictions about how object-based attention might manifest in neural data recording. The SpOC model postulates that an individual control channel – an individual spoke – exists for every possible form or potential target of attention. Attention to a region of space would be physically implemented as a channel originating in the attentional control areas and terminating on the early visual areas that receive their input from that region of the visual field.

How exactly activity in this control channel serves to enhance the processing or boost the representation of the visual information present in its targeted visual area remains an important unanswered question in the neuroscience of attention. One possible answer is that neural activity in the hub region mapping onto the target visual area simply increases the excitability of neurons in the target sensory site, boosting their signal gain. If a control channel exists for every form of attention and every attentional target, then an important implication is that there are individual channels terminating on visual areas that are primarily responsible for the processing of object categories. Thus, object-based attention, just like spatial attention and feature-based attention, is a distinct form of attention because the targets of its attentional modulation are object category-specific visual areas. Object-based attention should therefore be detectable in neural data as a distinct active subnetwork in the DAN. In a pure version of the SpOC model, object attention control signals are issued from the same central hub as those targeting other forms of visual information, and the channels over which the signals are conveyed and the mechanism of signal enhancement at the sensory sites are fundamentally the same in all those cases.

Whether this simple model accurately reflects how attention to different forms of visual information is carried out in the brain is a matter for further study – it is also conceivable that the channels over which control signals are sent are essentially different from each other and that the modes of enhancement differ across different attentional targets. Alternatively, the DAN might be exclusively a control system for modulating visual spatial inputs, with separate control networks involved in the modulation of sensory signals coding other visual attributes. In either case, the SpOC model would need to be modified. But the SpOC model in its simplest form, detailed above, generates testable predictions. For example, the same neural signatures of attention's effects in sensory areas should be observable in multiple types of attention tasks, when attention is directed toward different types of visual information. Specifically, some of the hallmarks of attentional gating in visual cortex during spatial attention tasks should also be observable during an object-based attention task.

Dynamic Gating by Controlled Neural Oscillations

As detailed in previous sections, a recurrent convolutional neural network is the anatomical architecture that supports the dynamic selectivity of sensory signals and abstract visual representations. Agnostic as to the ongoing validity of a source-site framework versus the potential merits of a new way of thinking (e.g., the Natural Object Attention Hypothesis), what remains is the question of how exactly sensory data is selected at early and mid-level processing stages, within the constraints of the recurrent neural network architecture. In the language of the SPoC model, the question is: What is the neural instantiation of the individual channels that connect source control areas, such as the dorsal attention network, to the sensory sites? And how does a signal from control areas, once it is issued over its specific channel, effect a modulatory change over its designated target neural representation? These mysteries boil down to a question of what the mechanism is that allows activity at high-level representations to interact with low-level representations. Because of the putative architecture of a recurrent neural network, direct activation over feedback connections in itself is not a sufficient explanation, because all upstream, low-level nodes connected to the high-level node would be simultaneously activated over recurrent channels in proportion to the strength of their connections. In a model architecture these connection strengths are numeric weights that are adjusted during learning, allowing the network to perform the task that it is trained for. But their direct equivalent in a real brain, synapses and the strength of synaptic connections, mediated by vesicle release, are anatomical and therefore not amenable to dynamic modulation on a time scale necessary for attention or adaptive behavior more generally.

One school of thought that has gained traction recently is that dynamic control of communication channels in brain networks is managed via the establishment, maintenance, and dispersal of oscillatory firing regimes. Neural oscillations can potentially solve the mystery of how dynamic patterns of activity can operate over neural architecture that is anatomically fixed, relative to the time scale of behavior and the ever-changing structure of the world around us.

What are neural oscillations?

In the early days of human EEG research, Hans Berger discovered that neural activity recorded at the scalp readily displays oscillatory behavior in the range of 8 – 13 Hz (Berger, 1929). These signals that Berger observed would appear most prominently as short-lived bursts of cyclical voltage changes over occipital scalp, especially when the subject of his recording held their eyes closed but remained awake. The sinusoidal waveforms were easily visible amidst the background EEG activity, even in live traces, because of their low frequency and high amplitude, relative to the rest of the EEG. Because of the prominence, consistency, and reliability of its oscillatory patterns, the 8 – 13 Hz range was named the alpha band. The 13 – 30 Hz range, also prominent but not to the same extent as the alpha band, was named the beta band. To this day, alpha band oscillatory waveforms remain the strongest electrophysiological signals measurable in the waking human brain (Jensen and Mazaheri, 2010), and the underlying physiology and functional role of this type of signal are matters of ongoing theory and investigation.

Oscillatory neural activity such as that seen in the alpha band is thought to reflect coordinated firing and cyclical patterns of excitability change across a widespread population of cortical neurons. The electrical activity recorded at the scalp with EEG almost always reflects moment-to-moment voltage changes arising from postsynaptic potentials in cortical pyramidal neurons, summated across large populations (Luck, 2014). For oscillations such as those in the alpha band to be visible in the recorded waveform of a scalp electrode, a vast population of spatially contiguous pyramidal neurons must be receiving synaptic input from upstream action potentials synchronously, so that their postsynaptic potentials, and their membrane excitability, are simultaneously and uniformly altered. The size of the neural population required to be acting synchronously for a typical alpha-band oscillatory pattern to appear at a scalp recording site is sufficient to potentially qualify the population as a computational module – an assembly of millions of neurons positioned in a larger neural network such that they play a discrete computational role.

Generally, slower oscillations reflect widespread synchronous membrane potential changes and faster oscillations reflect more spatially restricted synchronization (Buzsáki et al., 2013). The period of oscillation is a function of numerous biophysical and architectural properties of the generating neural network, including the limited speed of neural signal propagation due to conduction along the axon, the size of the neuronal pool engaged in a given cycle, the transmission length, the integration time for multiple upstream sources, and the membrane time constant of cortical pyramidal cells (Buzsaki and Draguhn, 2004; Buzśaki and Wang, 2012). Regardless of the size and spread of the neural population generating oscillatory activity, the interpretation that EEG oscillations reflect a precise underlying spatiotemporal consistency of spiking activity in neural assemblies hints to the functional significance of the underlying neural behavior.

Because of its prominence at occipital scalp sites in human EEG subjects during periods of drowsiness, mind-wandering, and closed eyes, alpha band activity has long been thought to reflect “cortical idling.” According to this interpretation, alpha band activity might arise from deactivated cortical regions, occurring in visual areas when they are not receiving any input or not engaged in processing sensory information (Pfurtscheller et al., 1996). More recently though, a positive functional role of alpha band activity has been suggested by experimental work and asserted by researchers: Instead of reflecting a cortical idling state that follows upstream cortical deactivation mechanisms, alpha band activity might be deliberately induced in cortex and thereby play a specific functional role (Klimesch et al., 2007; Jensen and Mazaheri, 2010). Thus, the frequency of activity belonging to the alpha signal might not just be an inconsequential emergent property or byproduct of cortical architecture and network electrophysiology when the network is disengaged from a larger information processing routine. Instead, it might reflect a specific computational mechanism and therefore an essential role in computation. Possible functional roles of alpha band activity will be discussed below.

Along with the alpha and beta bands, numerous other frequency bands of neural activity have been identified, named, and assigned conjectural roles in the brain's operation. The delta band refers to activity in the 1.5–4 Hz range; the theta band refers to activity in the 4–10 Hz range; the gamma band refers to activity in the 30–80 Hz range; the high gamma band refers to activity above 80 Hz; and various other slower and faster regimes have also been stratified and studied as discrete bands (Buzsáki et al., 2013). Altogether, oscillatory bands that have been identified cover five orders of magnitude in frequency.

Criticism of the explanatory invocation of neural oscillations

Gregory Hickok
@GregoryHickok

Ok. I'm gonna come out and say it. I think neural oscillations are the 2010s equivalent of mirror neurons. Real, but not doing all the work people think they are. I'm willing to be convinced otherwise but the more I look the more confident I get that there's not much there.

10:58 AM · Aug 12, 2020 · Twitter Web App

36 Retweets **15** Quote Tweets **280** Likes

Gregory Hickok @GregoryHickok · Aug 12
Replying to @GregoryHickok
[waits for backlash pile on...]

2 9

Gregory Hickok @GregoryHickok · Aug 12
Here's another way of expressing my doubt: pick any entrainment paradigm. In the middle of the entrainment at a random phase I say a word at 80 dB. Listener comprehends every time. What are the mechanisms that enable that entrainment override? Isn't that the real driver?

2 6

Gregory Hickok @GregoryHickok · Aug 12
Replying to @GregoryHickok
Btw, comparison to mirror neurons is not intended to denigrate. I find MNs important and interesting. I honed my thinking tremendously by testing and evaluating the range of hypotheses regarding their function. Some ideas turned out to be wrong, but that's how science works.

14

Figure 3. Screenshot of tweet by Dr. Gregory Hickock of University of California, Irvine, from August 12, 2020, suggesting that neural oscillations should not credited for as much of the brain's functioning as they seem to be nowadays. Accessed on January 19, 2021 at: https://twitter.com/GregoryHickok/status/1293607742054084608

As interest in the potential mechanistic roles of neural oscillations has accumulated, criticism has also mounted. There are two especially salient and related points of criticism.

The first point of criticism is that the range of cognitive and behavioral phenomena for which neural oscillations are invoked as an essential mechanism is implausibly large. For example, Buzsáki has pointed out that since the early observations of murine hippocampal theta oscillations in the 1930s, more than 50 different behavioral correlates have been proposed (Buzsáki, 2020). These behaviors include arousal, orienting, attention, volition, mismatch detection, movement, REM sleep, whisking, instrumental response, operant learning, voluntary movement, decision making, information processing, olfaction, anxiety, aggression, habituation, avoidance, play, hypnosis, working memory, plasticity, mapping, and navigation. If so many disparate behaviors are attributed to a specific form of neural activity – hippocampal theta oscillations – then either a large fraction of the research referred to is inaccurate in its interpretations, or an underlying cognitive or behavioral function is the true correlate of these oscillations, and it is scientifically imprecise to assert that the oscillatory activity is causally directly related to the behavior. Along with those linked to theta oscillations, similarly large lists of behavioral and cognitive correlates can be upturned for all the rest of the named frequency bands. The sheer number of cognitive and behavioral phenomena attributed to each named frequency band dilutes the sense of explanatory satisfaction accompanying the potential mechanistic role performed by each frequency band. In other words, the more putative distinct behavioral or cognitive correlates of a given form of oscillatory neural activity, the less likely it seems that any given one of those correlates is truly causally linked to the underlying neural activity in question.

The second point of criticism is that neural oscillations should not be used as an explanatory endpoint in studies of the neural activity underlying cognition and behavior. This second criticism is related to the first one in that one possible reason that so many disparate cognitive and behavioral phenomena are assigned mechanistic explanations that invoke the same type of neural oscillation is that the precise computational role that oscillatory neural activity, and how that role fits into a larger computational scheme, is not fully understood. Neural oscillations are readily detectable signals both at the level of single unit recording and in the summed electrical activity that can be recorded extracortically, but just because they are easy signals to identify with our current scientific methods does not make them fundamentally more important for the brain's functioning.

Rebuttal to criticism and proposed mechanisms

In summary, the major criticisms of neural oscillation science discussed above are that specific frequency bands of neural oscillations are invoked in explanations for an implausibly wide range of behaviors and computational processes, and that neural oscillations are too often relied upon as an explanatory recourse for numerous cognitive phenomena without sufficient understanding of how or why neural oscillations might play an important role therein. To the first criticism, proponents of the study of neural oscillations respond that because the field is nascent, the taxonomy of oscillatory behavior in the brain is poorly developed: Although the existing system of terminology assigns labels according to the frequency of neural activity, when a better understanding of the computational and mechanistic role of neural oscillations in various domains is attained, proper terminology will emerge that prioritizes role over frequency (Buzśaki and Wang, 2012; Buzsáki et al., 2013). The second point of criticism accurately describes the state of the art in the field of neural oscillations, but this criticism likely will lose its potency over time, because numerous lines of theory are being developed to attempt to explain why neural oscillations might be functionally significant, and evidence is accumulating to support those theories.

Neural oscillations exist both as rhythmic patterns of spiking activity that can be recorded in single units and as periodic voltage waveforms that reflect cyclically changing membrane excitability and are observable even extracortically. But if these oscillations are not epiphenomenal, the urgent question remains: What’s special about neural oscillations? Why are neural oscillations considered potential components of the brain’s computational processes?

One mathematical property of oscillators that may make them useful for neural computation is that synchrony between two oscillators can be easily achieved and sustained even if the connection between them is weak, and even if their intrinsic frequencies do not match precisely (Mirollo and Strogatz, 1990). In the brain, the oscillators in question might be neural assemblies or individual neurons. The connections between them would be synaptic, and the strength of the connection would be the magnitude of excitatory or inhibitory input onto the postsynaptic neuron and its likelihood of triggering an action potential. This mathematical property of coupled oscillators – the easy ability to form synchronous pairs – indicates how a dynamic system of linked neural populations can transiently induce synchronized cell assemblies among its constituents. Furthermore, because the dynamic system behavior in the brain naturally supports oscillatory behavior, it makes sense, evolutionarily, for the brain to take advantage of that behavior and utilize it in its computational processes. Neural oscillations are transiently stable regimes of excitability and spiking that can be easily and reliably established by small amounts of parameter tweaking. This property makes them less metabolically expensive and therefore more theoretically appealing than modulation of vesicle release kinetics as a means of dynamically altering computational networks to suit task demands.

One plausible reason that neural oscillations occur in the brain is that they are necessary for the transient binding of functional cell assemblies. According to this conjecture, distributed networks of neurons and neural groups are transiently brought into cooperation by oscillatory synchrony (Buzsaki and Draguhn, 2004). As detailed above, oscillatory synchrony is a metabolically cheap way to form temporary neuronal coalitions, compared to the energy-intensive biochemical steps involved in modulating vesicle release potentials at synapses according to transient computational demands. Once a temporary coalition of neurons is formed via oscillatory synchrony, a unique downstream population of neurons can be activated via the anatomical connections of the cooperating neural groups. Thus, in a sense, the distributed population brought into cooperation by oscillatory synchrony *codes* for the downstream target population in an activation pathway. And the computational complexity of the brain increases dramatically, because a different set of downstream activation targets maps onto every possible combination of neural groups that can participate in transient, synchronized coalitions. This theory of transient neural assembly binding suggests that neural oscillations play an indispensable functional role in the brain's operation, and that without the hierarchical system of oscillations, even the most rudimentary adaptive behaviors and cognitive functions would be impossible within the existing structure of the brain. Corroborating this line of thinking is the observation that the hierarchical system of oscillations is highly phylogenetically conserved across species with widely ranging brain volumes, lifespans, ecological niches, and other biological characteristics (Buzsáki et al., 2013).

As detailed above, theoretically, neural oscillations play an elemental role in the computational processes in the brain. But apart from theories about the plausibility of synchronous oscillators in the brain and their potential mechanistic role in the formation of transient networks, what further reason is there to believe that neural oscillations play a role in neural computation and cognition more broadly? Several theories about the mechanistic role of neural oscillations in cognition will be presented here. These theories are not mutually exclusive, and are each supported by empirical evidence.

Modulation of alpha band oscillatory activity is closely associated with selective attention, and is therefore thought to be a crucial component of the neural mechanisms underlying selective attention. When covert attention is directed to a location on one side of the visual field, alpha is more strongly suppressed over the visual hemisphere contralateral to the attended hemifield (Worden et al., 2000; Sauseng et al., 2005; Thut et al., 2006). This lateralized reduction of alpha activity observed in hemispatial covert attention tasks is thought to reflect an increase in cortical excitability in task-relevant sensory neurons: the early visual areas that receive their input from the regions of the visual field that are targeted by attention. It is speculated that oscillatory activity in the alpha band is deliberately reduced as a mechanism of attentional control, in order to facilitate the processing of upcoming stimulus inputs (Romei et al., 2008). Therefore, influential neural models of attention propose that top-down control signals from the DAN selectively modulate cortical excitability during attention via changes in alpha power. Although the findings that alpha band power decreases over brain areas that receive their input from attended spatial locations is correlational in nature, causal evidence in favor of this claim has also been observed. Repetitive transcranial magnetic stimulation (rTMS) showed disruptions in alpha lateralization after selectively disrupting processing in FEF and IPS (Capotosto et al., 2009, 2017), substantiating the hypothetical functional link between activity in the frontoparietal network and modulation of posterior alpha activity. And in a similar vein, single-pulse transcranial magnetic stimulation delivered to FEF, emulating attentional control signals endogenous to that region, has been shown to reset oscillatory phase in the beta frequency range over occipital sites, with accompanying perceptual modulation (Veniero et al., 2021).

One theory that seeks to explain why alpha band modulations are associated with shifts in selective attention is the Gating By Inhibition model. According to this model, the transmission of information through the brain is gated by systematic modulation of oscillatory activity in the alpha band (Jensen and Mazaheri, 2010). The membrane excitability of task-engaged brain areas is maintained through a process of lowering the power of activity in the alpha band, relative to disengaged areas. Gamma band activity, which is thought to reflect local neural processing, is inversely related to alpha band activity, such that an increase in local alpha band activity will result in a decrease in local gamma band activity, and hence a lowered capacity for neural computation or signal propagation within that local region. Thus the alpha band might represent a means of functionally inhibiting selected brain areas.

A related model, the Communication Through Coherence model (Fries, 2005), seeks to explain empirical findings that associate information processing and task engagement with coordinated local neural activity in the gamma range (Fries et al., 2007; Jensen et al., 2007). According to this model, the phase of the ongoing gamma rhythm in local neural activity reflects recurring temporal windows for optimal communication, and by synchronizing these windows of excitability, two neural populations will be able to communicate effectively and thereby propagate information. Neurons' intrinsic oscillations in the high frequency range associated with the gamma band represent periodic changes in membrane potential, and at the peaks of membrane excitability, a neuron is both more likely to initiate a spike, propagating a signal downstream, and to receive a signal by having its internal electrochemical state changed by upstream spikes. Thus, according to the Communication Through Coherence model, synchronized oscillatory activity between two neurons or between two neural populations effectively opens a channel for communication between them.

By extension, a top-down system that is able to selectively adjust the oscillatory power and phase of individual neural groups would be able to dynamically alter the pathways along which information is propagated through networks of brain areas.

The Gating By Inhibition model and the Communication Through Coherence model are linked by the observation that alpha band activity is inversely related to gamma band activity. In Gating By Inhibition, transient increases in alpha band activity causally suppress gamma band activity, thereby suppressing local information transmission, representation, and computation. The nature of the causal link between alpha band activity and gamma band activity has not been established and is a matter of ongoing investigation. However, there is another possible explanation for why top-down control over alpha band activity would affect the information capacity of a neural population and its ability to transmit information in its network that does not necessitate a causal relationship between alpha band activity and gamma band activity in order to support the theoretical involvement of alpha band modulation in attention.

Oscillatory neural activity, synchronized across a neural population, may reduce the information capacity of that population simply because of its periodic nature. From the perspective of information theory, information entropy is reduced as uncertainty is diminished, and thus the information conveyed by a time-varying signal is reduced in proportion to that signal's periodicity. The behavior of dynamical (time-varying) systems can be characterized along a spectrum from periodicity to chaos. Whereas periodic systems are predictable, systems with chaotic dynamics are exponentially sensitive to small perturbations. Theorists have proposed that computational capacity of dynamical systems is greatest when the system is near the phase transition between periodic and chaotic regimes. In this state of phase transition, known as criticality, the elementary components of computation – information representation, transmission, and modification – are thought to be maximally available, and modeling work has borne out this idea (Langton, 1990; Bertschinger and Natschläger, 2004; Boedecker et al., 2012). Altogether, this line of theoretical work suggests that the top-down induction of alpha band activity, which reflects widespread synchronized and periodic activity in a large neural population, might functionally inhibit the computational involvement of that neural population through periodicity itself. This possibility suggests a modification to Gating By Inhibition in which the gating of information processing effected by alpha band activity is not mediated through a causal connection to gamma band activity. The observation that gamma band activity is inversely related to alpha band activity may be only corollary, reflecting two separate underlying processes: top-down induction of alpha band synchronization to reduce computational capacity, and network routing via transient gamma band coalitions at the scale of local circuits.

Object-based attention and neural oscillations

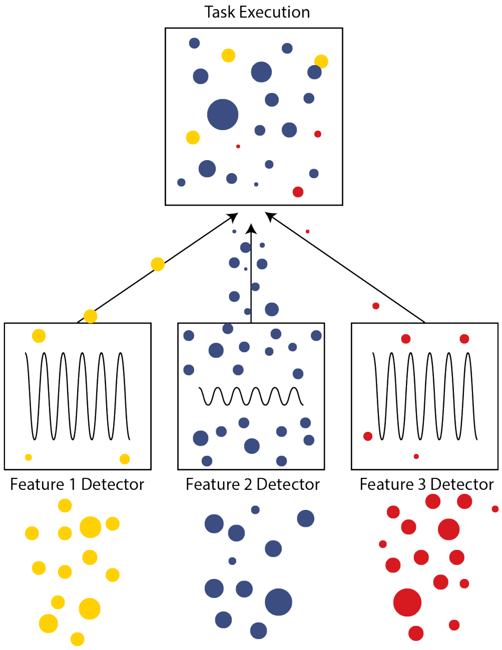


Figure 4. Illustration of the general tenets of a model of attention that describes the controlled flow of signal propagation by modulation of oscillatory neural activity.

Models of the brain's functional organization that highlight the role of oscillatory neural activity, such as the Gating By Inhibition model and the Communication Through Coherence model, propose that the pathways that guide the flow of information through the brain are fundamentally dynamic and gated by synchronized, oscillatory spiking activity and membrane excitability. By extension, these models indicate that signal propagation and information processing in the brain can be directed in a top-down manner by control signals that alter oscillatory firing patterns in different brain areas.

Because attention is generally conceived as a top-down control process whose function is to selectively enhance or suppress the information processing activity in targeted brain regions, such that targeted forms of sensory or internal information are elaborated or expedited toward higher level cognitive and behavioral domains, the dynamic gating of neural computation by controlled neural oscillations is an appealing hypothetical mechanism. The theory goes that top-down control areas, such as the DAN, can exert selective control over the activity of modular brain areas by altering the oscillatory firing patterns in those brain areas in a way that favors receptivity to upstream signals and conductance of signals downstream, while performing the converse function in task-irrelevant areas.

In terms of the Communication By Coherence model, top-down attentional control areas might synchronize the gamma band firing activity between an upstream region and a downstream region when that downstream region is necessary for computing the task-relevant properties of the sensory input. In line with prediction, this phenomenon has been observed in a study of top-down attention in monkeys, using intracranial LFP recordings (Bosman et al., 2012). Bosman and colleagues showed that when a monkey was instructed to attend selectively to the color of a visual display, gamma band activity synchronized between primary visual cortex and area V4, and did not synchronize between V1 and other visual areas that are thought to code for task-irrelevant visual information.

The Gating By Inhibition model suggests that top-down attention signals would increase alpha activity in task-irrelevant visual areas, relative to task-relevant visual areas. In visual attention research, this prediction has been supported in spatial attention tasks (Worden et al., 2000) and feature-based attention tasks (Snyder and

Foxe, 2010a), which have found relative focal alpha power decrements in task-relevant visual areas. Furthermore, a study by Liu and colleagues utilizing the simultaneous recording of EEG and fMRI has directly linked activity in the DAN to focal alpha modulation in visual areas during a cued spatial attention task in which on a trial by trial basis, covert attention was directed either to the left or right of a central fixation (Liu et al., 2016). The researchers correlated single-trial EEG alpha power in the post-cue period, from either hemisphere individually, with BOLD activity across the entire brain, to identify brain areas that covaried with alpha. To do this, they slightly modified the standard general linear model analysis method for fMRI BOLD activity. The standard method is essentially a test, at each voxel in the fMRI image, of how well the BOLD time course fits to a predicted function; that predicted function is a convolution of a canonical hemodynamic response function with a boxcar function that describes the timing of different stimulus events. But to test for coupling between single-trial alpha power and BOLD, the researchers modulated the height of the boxcars corresponding to the cue periods from individual trials, according to the relative degree of alpha band modulation in each trial. The results of the regression then specifically identified voxels in the fMRI data whose activity correlated closely with the degree of alpha band modulation that occurred on each trial. The most important result from this study was an inverse correlation between alpha power over the visual hemisphere contralateral to the cued visual hemifield and BOLD in the intraparietal sulcus, left, middle frontal gyrus, and inferior and middle temporal gyri. These areas have previously been identified as some of the areas in the dorsal attention network that is involved in top-down selective attention.

Within the conceptual framework laid out here, object-based attention is implemented by the same mechanism: Top-down control signals from regions like the DAN selectively modulate alpha band power to enhance the receptivity and conductivity of neural populations whose networked activity codes for an attended object category. Further chapters of this dissertation elborate an empirical exploration of this prediction.

Conclusion and Future Directions

In the present chapter, I aimed to provide an overview of an exciting frontier in the cognitive neuroscience of attention: how attention operates in real world settings. This question is vital to the progress of cognitive neuroscience, because attention is primarily conceptualized by virtue of its role in natural behavior. I used the operationalized phenomenon of object-based attention as a way to broach the topic of real-world attention.

I began this chapter by describing early thinking about attention that emphasized the role of attention in natural settings. I provided a general overview of attention research in cognitive neuroscience, including the ways that it has been operationalized for empirical study, and foundational findings that point to its neural mechanisms in the brain. I then shifted to the topic of how the brain represents objects, to contextualize the topic of object-based attention. I reviewed a new school of thought that object-based attention may be the natural mode of attention in the brain, and described studies that investigated how object-based attention may be instantiated in the visual system, especially in light of the latest understanding of how objects are represented.

My synthesis of object recognition theory and object-based attention theory included a proposal for a new view of attention control in the brain that is based primarily on the findings that objects may be the primary units of attention, and that objects are represented with a distributed population code embedded in a deep, recurrent, convolutional neural network whose essential function is to extract successively abstract information about visual input and embed statistical regularities of the environment into its network structure. This new conceptualization of selective attention presents an alternative framework to the conventional site-source dichotomy of attention, epitomized by the SpOC model, which was also presented in this chapter. I called the new proposal about attention’s neural organization the Natural Object Attention Hypothesis in order to convey the intimate relationship between object representation and attention in the brain, and to emphasize that by embedding the statistical properties of natural sensory input, the brain may not only be encoding representations of statistical regularities that we refer to as objects, but that momentary task demands that selectively activate object representations, a phenomenon that we call attention, might actually be a further abstract statistical property of sensory input.

The Natural Object Attention Hypothesis is dependent on the understanding that the visual system, perhaps complexed with frontal and parietal areas constituting the dorsal attention network, is a recurrent deep convolutional network. Thus artificial recurrent deep CNNs can be used as a test bed for research into this hypothesis, and object-based attention more generally. Hypothesized attentional mechanisms, such as those of the Buschman-Kastner theory (Buschman and Kastner, 2016), can be tested in artificial CNNs to better understand how broadly distributed attentional control signals can be refined into selective modulation of sensory information via recurrent connections. Furthermore, artificial recurrent CNNs can be used to test predictions of the Natural Object Attention Hypothesis. If we trained recurrent CNNs to classify tasks involving object images rather than classify object images, would we find similar patterns of object category-specific activation in both higher layers and lower layers, as might be predicted on the basis of human studies (Peelen and Kastner, 2011)?

Another question that deserves future research to strengthen the object-based attention school of thought pertains to the dramatic discrepancy between simple conjunction search and object search in natural scenes. As described in a previous section of this manuscript, on the basis of results from classical search paradigms, it may be predicted that real world search would take an extremely long time. The number of items in the “search array” of a real-world scene is extremely large and varied. The target is conceptually distinct from the distractor items but in terms of image properties, the target and distractors may be very similar and distractors are as dissimilar from each other as they are from the target. We intuitively know that real-world search is not as difficult as might be predicted from classical search findings, and laboratory findings using more ecologically valid search designs confirm that ecological search can appear to be almost parallel in its performance. So how can we make sense of classical search paradigms? Are the designs of classical attention paradigms just special cases of more ecologically valid, real-world attention?

The question of whether the findings and models of classical attention theory can be subsumed by a more general understanding, and with it the implicit question of what the value is of the classical way of thinking about attention, hints at a much larger question about the nature of science. All scientific models are simplifications of the real phenomena that they depict, and so necessarily they have limited applicability. Generally, the best model is not the one that contains the most detail about its subject, because by that criterion the only good model of the brain would be the brain itself. Instead, a model is good simply when it is clearly stated, descriptive, and predictive. Its essential character can be captured in explicitly defined factors and variables that may lend themselves to mathematical formulation. Its factors are useful in that they describe the character of the phenomenon in question, such that the phenomenon can be recognized from the description, and any other detail would be superfluous. And application of the model makes predictions that can be empirically tested to judge its explanatory value. According to these criteria, parceling the function of the brain into discrete cognitive functions, including attention and object recognition, makes sense. Predictions made about the behavior of the brain within the frameworks modeling these discrete phenomena should be expected to have gaps and inaccuracies when extended into new domains. But when the inaccuracy of models and the limits of their explanatory power hinder the progress of a scientific field, it should be taken as a sign that the level of explanation entailed by those models is inappropriate for that field’s overarching task.

In cognitive neuroscience, the overarching task is to draw the biology of the brain closer to the subjective phenomena of psychology. Attention is a fundamental explanandum of psychology, so the explanatory limitations imposed by classical attention research suggest that a new level of description is necessary. In this chapter, I have proposed the Natural Object Attention Hypothesis as a step towards a new model of brain function that accommodates the useful terminology of attention and object recognition, but does not entangle theory with paradox. Future work to develop and test predictions of this new framework, and absorb previous theory into the new model, will advance the field of cognitive neuroscience toward its goal of merging mechanistic biological processes with the intangible subjective qualities of human nature. Until those days when such work is possible, the SpOC model, and other frameworks for understanding attention that utilize a site-source dichotomy, will continue to advance our understanding of attention's neural mechanisms, and drive the empirical wedge further into the profound mysteries of human psychology.

Chapter 2: Experiment 1

Introduction

Selective attention is a fundamental cognitive ability that facilitates the processing of task-relevant perceptual information and suppresses distracting signals. The influence of attention on perception has been demonstrated in improvements in behavioral performance (Posner, 1980) and changes in psychophysical tuning curves (Carrasco and Barbot, 2019). In humans, these perceptual benefits for attended stimuli co-occur with enhanced sensory evoked potentials (Van Voorhis and Hillyard, 1977; Eason, 1981; Mangun and Hillyard, 1991; Eimer, 1996; Luck et al., 2000) and increased hemodynamic responses (Corbetta et al., 1990; Heinze et al., 1994; Mangun et al., 1998; Tootell et al., 1998; Martínez et al., 1999; Hopfinger et al., 2000; Giesbrecht et al., 2003). In animals, electrophysiological recordings indicate that sensory neurons responsive to attended stimuli have higher firing rates than those of unattended stimuli (Moran and Desimone, 1985; Luck et al., 1997), improved signal-to-noise in information transmission (Mitchell et al., 2009; Briggs et al., 2013), and increased oscillatory responses (Fries et al., 2001) that support higher interareal functional connectivity (Bosman et al., 2012).

Most models of selective attention posit that top-down attentional control signals arising in higher level cortical networks bias processing in sensory systems (Nobre et al., 1997; Kastner et al., 1999; Corbetta et al., 2000; Hopfinger et al., 2000; Corbetta and Shulman, 2002; Petersen and Posner, 2012). However, precisely how top-down signals influence sensory processing within sensory cortex remains unclear. One possible mechanism involves the modulation of EEG alpha oscillations (8 – 12 Hz). When covert attention is directed to one side of the visual field, alpha is more strongly suppressed over the contralateral hemisphere (Worden et al., 2000; Sauseng et al., 2005; Thut et al., 2006; Rajagovindan and Ding, 2011). This lateralized alpha reduction is thought to reflect an increase in cortical excitability in task-relevant sensory neurons in order to facilitate the processing of upcoming stimuli (Romei et al., 2008; Jensen and Mazaheri, 2010; Klimesch, 2012). A link between top-down activity in the frontal-parietal attentional control system and alpha in sensory cortex has been suggested by studies using transcranial magnetic stimulation to control regions (Capotosto et al., 2009, 2017), simultaneous EEG-fMRI studies (Zumer et al., 2014; Liu et al., 2016) and magnetoencephalography (Popov et al., 2017).

Although the majority of studies of the role of alpha in selective visual attention have focused on spatial attention, alpha mechanisms may be more general (Jensen and Mazaheri, 2010). Selective attention to low level visual features – motion versus color – has also been shown to modulate alpha that was localized to areas MT and V4 using EEG modeling in humans (Snyder and Foxe, 2010a). Therefore, it appears that attention-related alpha modulation can occur at multiple early sensory processing levels in the visual system, with the locus of alpha modulation functionally corresponding to the type of visual information being targeted by attention. It is unknown whether the alpha mechanism is also involved in attentional control over higher levels of cortical visual processing, such as attention to objects. In the present study, we tested the hypothesis that alpha modulation is a mechanism for selective attention to objects by recording EEG from participants performing an anticipatory object attention task using three categories of objects: faces, scenes, and tools. Using EEG decoding methods we provide support for this hypothesis by revealing object-specific modulations of alpha during anticipatory attention to different object categories.

Materials and Method

Participants

EEG data were recorded from 22 healthy undergraduate volunteers at the University of California, Davis. All participants had normal or corrected-to-normal vision, gave informed consent, and received course credit for their participation. Two volunteers opted to discontinue their participation midway through the experiment; data from the remaining 20 participants were used for all analyses.

Apparatus and stimuli

Participants were comfortably seated in an electrically-shielded, sound-attenuating room (ETS-Lindgren, USA). Stimuli were presented on a VIEWPixx/EEG LED monitor, model VPX-VPX-2006A (VPixx Technologies Inc., Quebec Canada), at a viewing distance of 85 cm, vertically centered at eye level. The display measured 23.6 inches diagonally, with a native resolution of 1920 by 1080 pixels and a refresh rate of 120Hz. The recording room and objects in the room were painted black to avoid reflected light, and it was dimly illuminated using DC lights.

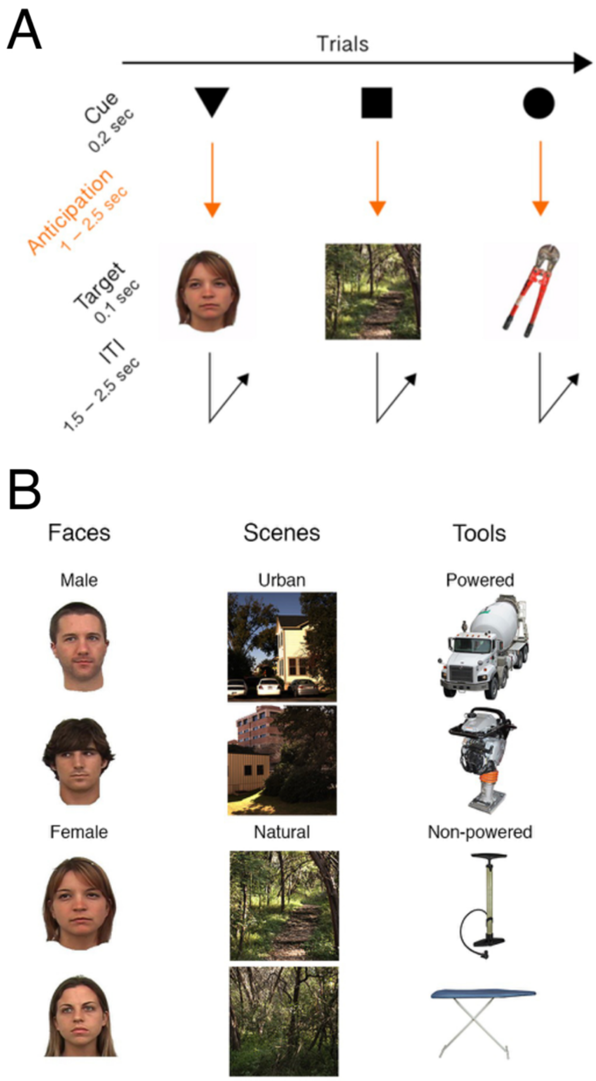


Figure 1. A. Example trial sequence for the attention task. Each trial began with the presentation of a symbolic cue that the subjects were taught predicted (80%) a specific object category. Following an anticipation period (cue-to-target) varying from 1.0 to 2.5 s, a picture of an object (face, scene or tool) was presented. On 20% of the trials one of the two uncued targets pictures were presented. Subjects were required to make a rapid-accurate discrimination of aspects of the pictures in both the expected and unexpected conditions (see text for details). **B. Examples of target images presented in the attention task.** Face, scene and tool pictures were selected from online databases.

Each trial began with the pseudorandomly selected presentation of one of three possible cue types for 200 msec (1° x 1° triangle, square, or circle, using PsychToolbox;

Brainard, 1997; Figure 1A). Valid cues informed participants which target object category (face, scene, or tool, respectively) was likely to subsequently appear (80% probability). Cues were presented 1° above the central fixation point. Following pseudorandomly selected SOAs (1000 – 2500 msec) from cue onset, target stimuli (5° x 5° square image) were presented at fixation for 100 msec. On a random 20% of trials, the cues were invalid, incorrectly informing participants about the upcoming target object category. For these invalid trials, the target image was drawn with equal probability from either of the two non-cued object categories. All stimuli were presented against a gray background. A white fixation dot was continuously present in the center of the display.

Target images (Figure 1B) were selected from 60 possible images for each object category. All target images were gathered from the Internet. Face images were front-face, neutral-expression, white-ethnicity faces, cropped and placed against a white background (Righi et al., 2012). Full-frame scene images were drawn from the University of Texas at Austin's natural scene collection (Geisler and Perry, 2011) and campus scene collection (Burge and Geisler, 2011). Tool images, cropped, and placed against a white background, were drawn from the Bank of Standardized Stimuli (Brodeur, Mathieu B.; Guerard, Katherine; Bouras, 2014). A pseudorandomly distributed inter-trial-interval (ITI; 1500 – 2500 msec) separated target offset from the cue onset of the next trial. Each set of 60 object images comprised 30 images of two different subcategories: male/female faces, urban/natural scenes, and powered/non-powered tools.

Procedure

Participants were instructed to maintain fixation on the center of the screen during each trial, and to anticipate the cued object category until the target image appeared. They were further instructed to indicate the target image object subcategory (e.g., male/female) with a button press as quickly and accurately as possible upon target presentation, using the index finger button for male (face), nature (scene), and powered (tool), and to press the middle finger button for female (face), urban (scene), and non-powered (tool). Responses were only recorded during the ITI between target onset and the next trial. Trials were classified as correct when the recorded response matched the target image subcategory, and incorrect when the response did not match, or when there was no recorded response. Each experiment block included 42 trials, lasting approximately 3 mins. Each participant completed 10 blocks of the experiment.

EEG recording

Raw EEG data were acquired with a 64-channel Brain Products actiCAP active electrode system (Brain Products GmbH), and digitized using a Neuroscan SynAmps2 input board and amplifier (Compumedics USA, Inc.). Signals were recorded with Scan 4.5 acquisition software (Compumedics USA, Inc.) at a sampling rate of 1000 Hz and a DC to 200 Hz online band pass. Sixty-four Ag/AgCl active electrodes were placed in fitted elastic caps using the following montage, in accordance with the international 10-10 system (Jurcak et al., 2007): FP1, FP2, AF7, AF3, AFz, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT9, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, FT10, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP9, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, TP10, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, PO9, O1, Oz, O2, PO10; with channels AFz and FCz assigned as ground and online reference respectively. Additionally, electrodes at sites TP9 and TP10 were placed directly on the left and right mastoids. The Cz electrode was oriented to the vertex of each participant's head by measuring anterior to posterior from nasion to inion, and right to left between preauricular points. High viscosity electrolyte gel was administered at each electrode site to facilitate conduction between electrode and scalp, and impedance values were kept below 25 kΩ. Continuous data were saved in individual files corresponding to each trial block of the stimulus paradigm.

EEG preprocessing

All data preprocessing procedures were completed with the EEGLAB Matlab toolbox (Delorme and Makeig, 2004). For each participant, all EEG data files were merged into a single dataset before data processing. Each dataset was visually inspected for the presence of bad channels, but no such channels were observed. The data were Hamming window sinc FIR filtered (1 – 83 Hz), and then down sampled to 250 Hz. Data were algebraically re-referenced to the average of all electrodes, and then further low-pass filtered to 40 Hz. Data were epoched from 500 msec before cue onset to 1000 msec after cue onset, so that anticipatory data from all trials could be examined together. Data were visually inspected to flag and reject trials with muscle tension artifact and eye movement artifacts that occurred during cue presentation. Independent component analysis (ICA) decomposition was then used to remove artifacts associated with blinks and eye movements.

EEG analysis

We used a power spectral density procedure, with the Matlab *periodogram()* function (window length 500 msec, step length 40 msec), to extract alpha band power for each electrode, for each participant and cue condition. Within each participant and cue condition, power spectral density results were averaged across trials. Averaged power spectral density results were used to visually examine alpha band power topographies across cue conditions.

We implemented a decoding analysis to quantitatively assess whether object attention was systematically associated with changes in phase-independent alpha band power topography across conditions. This analysis routine was adapted from a routine to decode working memory representations from scalp EEG (Bae and Luck, 2018).

Decoding was performed independently at each time point within the epochs. We implemented our decoding model with the Matlab *fitecoc()* function to use the combination of a support vector machine (SVM) and error-correcting output coding (ECOC) algorithms. A separate binary classifier was trained for each cue condition, using a one-versus-one approach, with classifier performance combined under the ECOC approach. Thus, decoding was considered correct when the classifier correctly determined the cue condition from among the three possible cue conditions, and chance performance was set at 33.33% (one-third).

The decoding for each time point followed a six-fold cross-validation procedure. Data from five-sixths of the trials, randomly selected, were used to train the classifier with the correct labeling. The remaining one-sixth of the trials were used to test the classifier, using the Matlab *predict()* function. This entire training and testing procedure was iterated 10 times, with new training and testing data assigned randomly in each iteration. For each cue condition, each participant, and each time point, decoding accuracy was calculated by summing the number of correct labelings across trials and iterations, and dividing by the total number of labelings.

We averaged together the decoding results for all 10 iterations to examine decoding accuracy across participants, at every time point in the epoch. At any given time point, above-chance decoding accuracy suggests that alpha topography contains information about the attended object category. However, a comparison of decoding accuracy to chance, by itself, is not sufficient for assessing whether an inference made on the basis of decoding accuracy is reliable. Although a one-way *t*-test of decoding accuracies across subjects against chance would provide a *t*-value and a statistical significance result for the time point in question, conducting the same test at each of the 375 time points included in our epoch would require a correction for multiple corrections that would result in overly conservative statistical tests. Therefore, following Bae and Luck (2018), we utilized a Monte Carlo simulation-based significance assessment to reveal statistically significant clusters of decoding accuracies.

By the Monte Carlo statistical method, decoding accuracy was assessed against a randomly chosen integer (1, 2, or 3), representing an experimental condition, for each time point. A *t*-test of classification accuracy across participants against chance was performed at each time point for the shuffled data. Clusters of consecutive time points with decoding accuracies determined to be statistically significant by *t*-test were identified, and a cluster *t*-mass was calculated for each cluster by summing the *t*-values given by each constituent *t*-test. Each cluster *t*-mass was saved. This procedure was iterated 1000 times, to generate a distribution of *t* masses to represent the null hypothesis that a given cluster of *t*-masses from our decoding analysis was likely to have been found by random chance. The 95% cutoff *t*-mass value was determined from the permutation-based null distribution and used as the cutoff against which cluster *t*-masses calculated from our original decoding data could be compared. Clusters of consecutive time points in the original decoding results with *t*-masses exceeding the permutation-based threshold were deemed statistically significant.

We performed the same decoding routine on phase-independent EEG oscillatory activity in the theta range (4 – 7 Hz), beta range (16 – 31 Hz), and the gamma range (32 – 40 Hz) to test the hypothesis that object attention-based modulations of EEG activity are specific to the alpha range. For filtering EEG data to the beta and gamma band, we set the minimum filter order to be three times the number of samples in the experimental epoch. For filtering to the theta band, we set the minimum filter order to be two times the number of samples, because the duration of the epoch was not long enough to allow a filter order three times the number of samples.

Results

Behavioral results

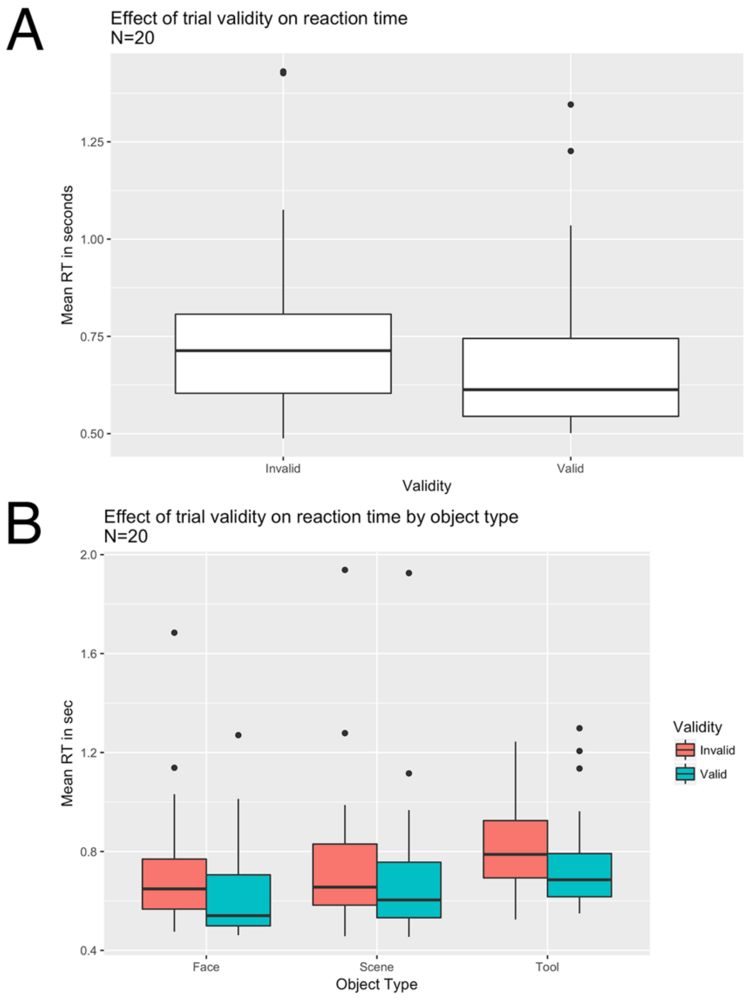


Figure 2. Behavioral Measures of Attention in Experiment 1. A. Box plots of reaction time data for invalid and valid trials for 20 subjects, averaged across attention (object) conditions. Thick horizontal lines inside boxes represent median values. First and third quartiles are shown as lower and upper box edges. Vertical lines extend to most extreme data points excluding outliers. Dots above plots represent outliers, defined as any value greater than the third quartile plus 1.5 times the interquartile range. Subjects were significantly faster overall for cued (valid) objects than uncued (invalid) objects. **B.** Reaction times for valid and invalid trials separately for each attention condition. Subjects were significantly faster for cued (valid) objects than uncued (invalid) objects for each object category.

Observed response accuracies were high and uniform across all object conditions and validity conditions (Invalid Face 96.6%, Invalid Scene 97.1%, Valid Face 96.8%, Valid Scene 96.7%, Valid Tool 93.1%) with the exception of the invalid attend-tool condition (87.5%), which we address below.

To determine whether our task elicited a behavioral attention effect, we compared reaction time (RT) for target discriminations between validly and invalidly cued trials. We observed faster mean RTs for valid trials than for invalid trials, averaging across conditions (Figure 2A) and for each condition separately (Figure 2B).

To quantitatively assess the effect of cue validity on RT, we fit a gamma-distributed generalized linear mixed model to the RT data (Lo and Andrews, 2015). We found a significant effect of *Validity* (valid vs. invalid; $p < 0.001$), with an estimated difference between valid and invalid trials of 68 msec. The model also revealed a significant main effect of object category ($p < 0.001$), due to the slower overall reaction times in the tool category. Thus, subjects were less accurate and slower in their responses to the tool category. Despite these slight performance decrements for the tool category, there was nonetheless a significant behavioral attention effect for the tool category, providing evidence that the subjects utilized all three cues types to prepare to discriminate and respond to the upcoming objects.

Alpha topography results

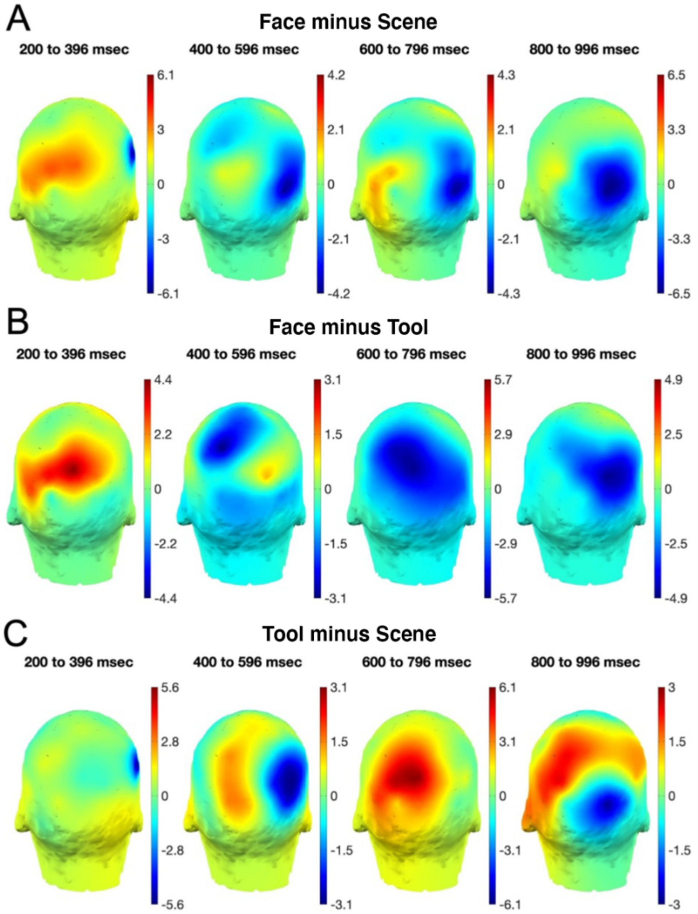


Figure 3. Topographic Difference Maps for Alpha Power in Experiment 1. A. Difference Maps for Anticipatory Attention to Faces minus Scenes**.** Alpha topography difference plot for attend-face minus attend-scene condition, averaged over participants, for four time windows relative to cue onset. The topographic difference maps are only shown until 1000 msec after cue onset, when the shortest latency targets could appear. The view of these difference maps if from behind the head. See text for description. **B.** Difference Maps for Anticipatory Attention to Faces minus Tools. Alpha topography difference plot for attend-face minus attend-tool condition, averaged over participants. **C.** Difference Maps for Anticipatory Attention to Tools minus Scenes. Alpha topography difference plot for attend-tool minus attend-scene condition, averaged over participants.

To *qualitatively* assess whether the pattern of alpha power across electrodes was different for anticipatory attention to the three cued categories of objects, we inspected topographic plots of alpha power for each object condition at different time periods following the cues, but prior to the onset of the target stimuli. In order to highlight differences between the alpha topographies between conditions, and to control for non-specific effects of behavioral arousal, we created pairwise alpha topography difference maps of one object attention condition subtracted from another object attention condition.

We observed that differences in alpha topography between object conditions emerged and evolved over the anticipatory (cue-to-target) period (Figure 3). In the attend-face minus attend-scene topographies (Figure 3A), we observed increased alpha power over the left posterior scalp, and decreased over the right posterior scalp during the course of the anticipatory period, with the lateralization becoming most prominent at longer post-cue latencies. In the attend-face minus attend-tool topographies (Figure 3B), the pattern was similar at the longest latencies, but more variable in intermediate periods of time. In the attend-tool minus attend-scene topographies (Figure 3C), the pattern of alpha differences was distinctive from those involving attend-face conditions; at the longest post-cue latencies the pattern of alpha power over the scalp was reversed from that in the other difference maps, with alpha power being higher over the left than the right posterior scalp. Overall, the presence of these difference among conditions is consistent with variations in the underlying patterns of cortical alpha power during anticipatory attention to faces, scenes, and tools. However, given the variability across subjects, and the inherent difficulty in quantifying difference maps between subjects across attention conditions, we turned to the method of EEG decoding to quantify the differences in alpha power across the conditions that are qualitatively described in the foregoing.

SVM decoding results

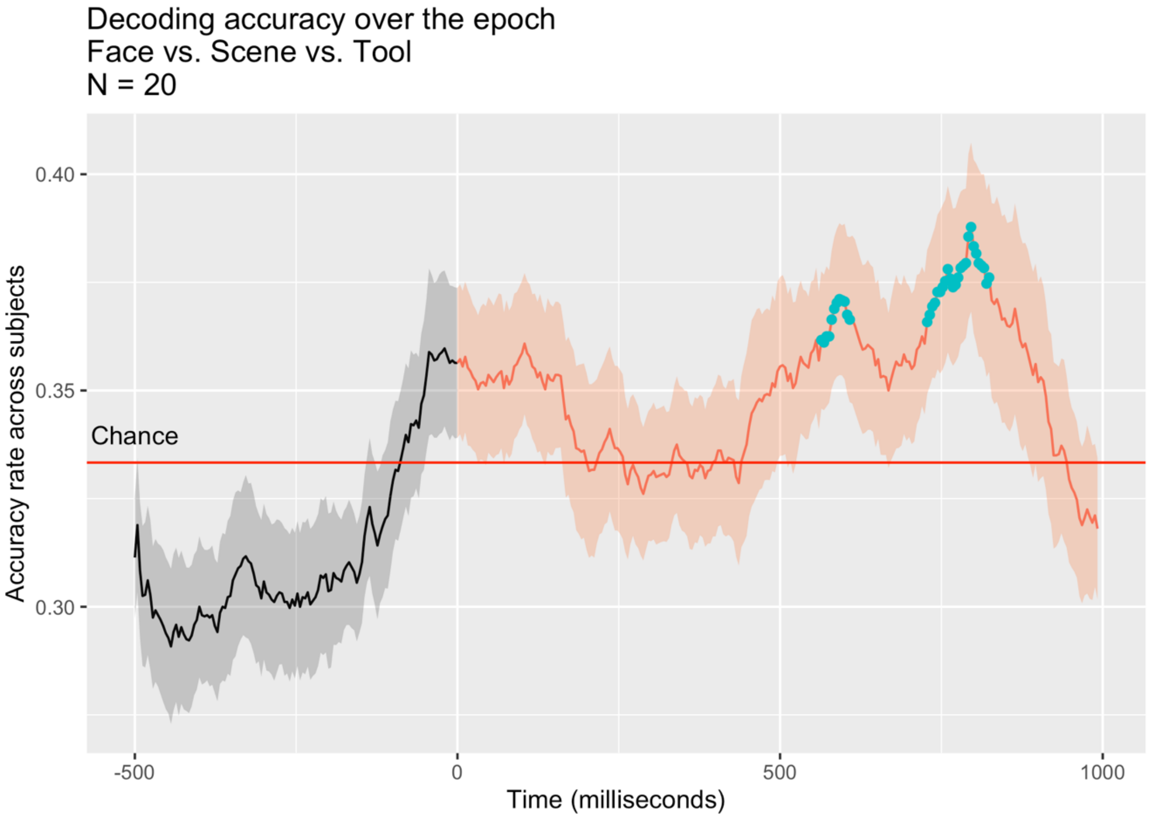


Figure 4. Alpha Band Decoding Accuracy for Experiment 1. Decoding accuracy of alpha band activity over the epoch, across participants. The horizontal red line represents chance decoding accuracy. The solid time-varying line is the across-subject mean decoding accuracy at each time point, and the shaded area around this line is the standard error of the mean. The grey shading denotes the pre-cue period, and the orange shaded segment represents the anticipatory period between cue onset (0 msec) and the earliest target onset (1000 msec). The turquoise dots denote time points that belong to statistically significant clusters of decoding accuracy, as determined by Monte Carlo assessment.

SVM decoding results (Figure 4) revealed statistically significant decoding accuracies in two clusters of time points around the range of 500 – 800 msec post-cue and pre-target (turquoise dots in Figure 4). Decoding accuracies in the range of -100 to +200 msec around the onset of the cue did not reach the threshold for statistical significance.

SVM decoding results for beta and gamma band oscillatory EEG activity revealed no statistically significant decoding in the anticipatory period (Figure 5). This result was in line with previous SVM decoding of alpha power topography in selective attention tasks.

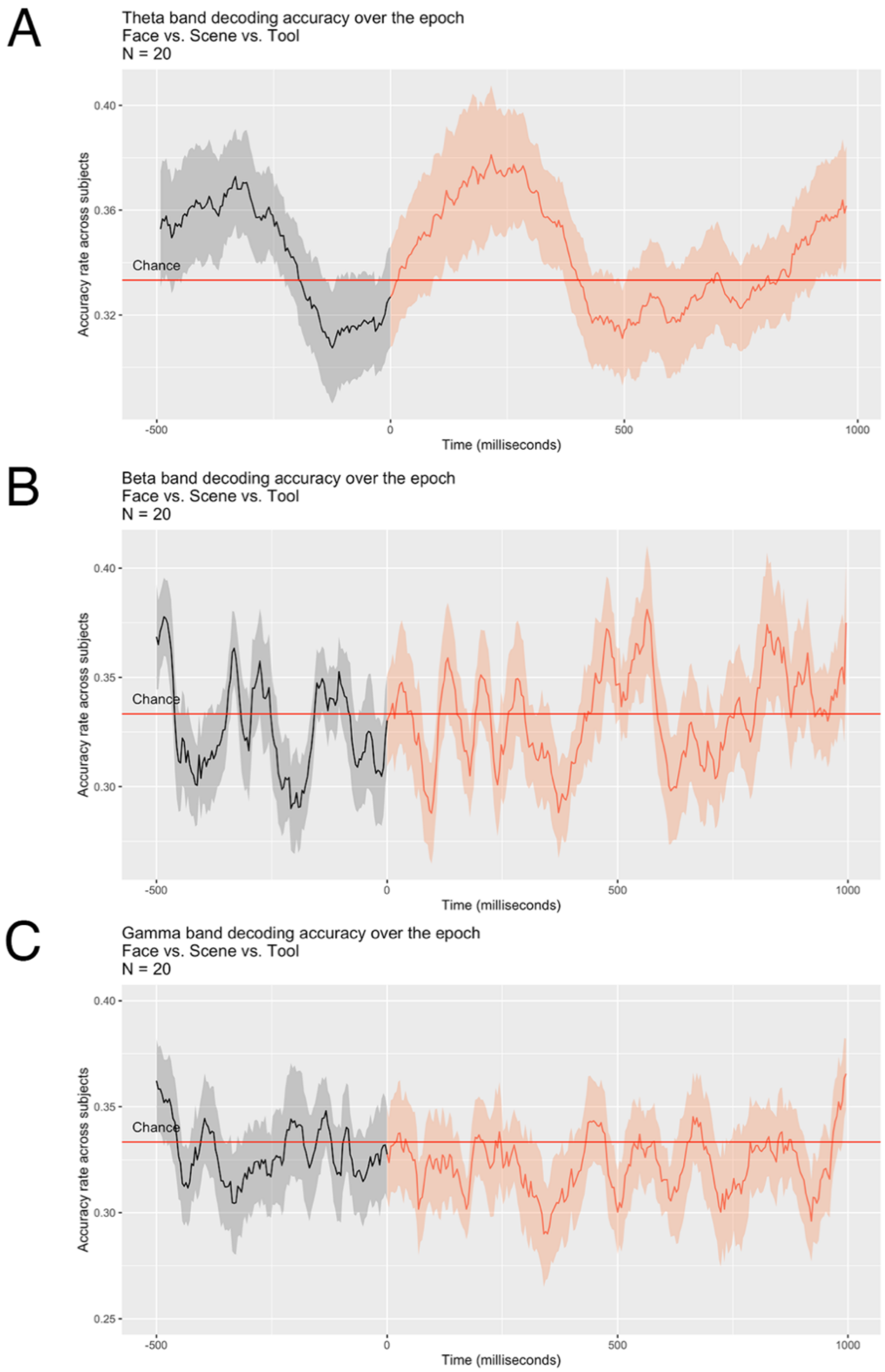


Figure 5. Decoding for Different EEG Frequency Bands in Experiment 1. A. The same SVM decoding procedure and Monte Carlo statistical procedure that was used for analyzing the alpha band data was applied to the theta band (4 – 7 Hz). **B.** The same decoding pipeline was applied to the beta band (16 – 31 Hz), revealing no statistically significant clusters of above-chance decoding accuracy in the preparatory period. **C.** The same decoding pipeline was also applied to the gamma band (32 – 40 Hz), and similarly revealed no statistically significant clusters of above-chance decoding accuracy in the preparatory period.

Chapter 3: Control Experiments

The results of Experiment 1 suggested that alpha power topographies systematically varied across the three object attention conditions. Behavioral results indicated that the participants of Experiment 1 were engaging a form of anticipatory attention that targeted object categories, guided by symbolic shape cues with learned meanings. On the basis of the finding that alpha power topographies differed systematically across the object attention conditions, we concluded that alpha band modulation occurs differentially within the cortex depending on the object category being attended, supporting a hypothesis that alpha band modulation is an essential component of the neural mechanisms of selective visual attention throughout the visual system. However, the design of Experiment 1 was imperfect in several ways, limiting the extent of our confidence in our theoretical interpretation. In order to address these limitations, we executed control experiments that were designed specifically to correct imperfections in the design of Experiment 1 and support our conclusions.

Pilot Experiment

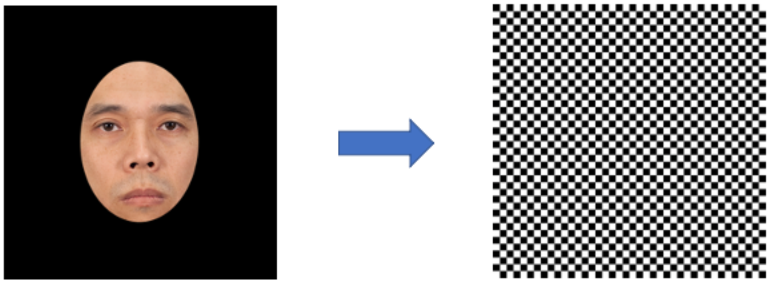
The experiment that we designed immediately after completing Experiment 1 never made it past the pilot stage, but for the sake of comprehensiveness the details of its design and limited behavioral results will be presented in this section. Because we terminated data collection for this experiment after finishing the first round of pilot data collection and subsequently moved onto a new control experiment design, this aborted pilot experiment will be referred to throughout this manuscript as the Pilot Experiment, rather than designating it as a full experiment with an ordinal number.

The rationale for the Pilot Experiment was to address the possibility that attention was not necessary for the successful execution of the task in Experiment 1. Despite the behavioral results we obtained indicating that Experiment 1 was successfully inducing an object-based attention effect, following the logic of operationalizing anticipatory attention by using an instructional cue and a valid/invalid trial design, we sought to strengthen the credibility of our claim by replicating the results of Experiment 1 with a new design in which attention is more clearly necessary to perform the task. In Experiment 1, the participants potentially could have largely ignored the instructional cue and waited until target presentation to engage task performance. To give a concrete example: A participant could have ignored the instructional cue at the beginning of a trial, not prepared any kind of task set or engaged anticipatory object-based attention to the cued object category during the cue-target interval, waited for the target image to appear, and when it appeared, quickly identify the object category, recall the button-mapping instructions for that object category, identified the subcategory and respond with a button press accordingly. This alternative method of task performance would be possible because the participants were trained in advance of the data collection period about the three object categories, their subcategories, and the button mapping for each object category.

In theory, if participants were utilizing this alternative task execution strategy, we would not have observed any difference between validly cued and invalidly cued trials, because the cues would have been completely ignored by the participants and thus there would be no systematic difference between those conditions. In Experiment 1 we did observe a statistically significant difference in behavioral measures between validly cued and invalidly cued trials, and moreover, the direction of the effect was as expected: Invalidly cued trials begot longer reaction times and lower accuracies than validly cued trials, in accordance with the standard pattern observed when attention is operationalized with a cuing paradigm. However, it remained a possibility that some participants performed the task without invoking object-based attention, or that participants sometimes utilized the instructional cues as intended but sometimes ignored them, and that the alpha power results were contaminated by trials or participants failing to comply with the task instructions. In order to preclude this possibility from invalidating our interpretation of the results from Experiment 1, we designed the Pilot Experiment in a way that would require object-based attention in order to be performed at all.

In Experiment 1, all target images were displayed on screen for 100 msec. Stimulus duration of this length is sufficient for easy perception. In the Pilot Experiment, we sought to reduce the duration of the target stimulus below the perceptual threshold for each individual participant. Shortening the target stimulus duration in this way would require the allocation of anticipatory attention to the target to improve target perception enough for the task to be performed. According to our design, without attention to the target, the image would not be perceptible, and thus the task would be impossible to perform. This scenario would be invoked on invalidly cued trials, when participants were cued to expect an object category different from the one that eventually appeared in the target stimulus. Our behavioral measures would be indistinguishable from chance across these invalid trials, whereas across trials where the object category was validly cued, perception would be improved enough for performance to be significantly greater than chance. Therefore, we still operationalized attention with a cued attention design and an invalid/valid trial dichotomy, but unlike Experiment 1, our criterion for the presence of attention would not simply be a difference in reaction time and accuracy in the expected direction, but a significant difference between observed behavioral measures and chance level in the valid trials, and no such significant difference in the invalid trials.

In order to make the target stimuli difficult to perceive and necessitate object-based attention, in addition to shortening the stimulus duration time to perceptual threshold on an individual participant basis, we overlaid each image with a random noise pattern. Moreover, after target stimulus offset, a checkerboard mask immediately was displayed, so that the identification task could not be performed on the basis of after image or iconic memory. The perceptual threshold target stimulus duration was found for each participant by adjusting it manually during the training phase of the experiment, before data collection, until a classification accuracy of 50% was reached.



~25 milliseconds (varying across participants)

1500 – 2500 msec

Figure 1. In the Pilot Experiment, the target image was displayed on screen for an amount of time that was found to correspond to the perceptual threshold of the image, and immediately followed by a checkerboard mask.

The Pilot Experiment also allowed us to remediate several limitations of the design of Experiment 1 that were unrelated to the central concern just discussed. We excluded the scene and tool object categories from the stimulus design and included letters as a new object category. Letters were chosen as the new object category because in the brain, letter representation is thought to be the contra-hemispheric complement of face representation in the visual cortex. This hemispheric lateralization of object representation in our task would grant us the greatest chance of localizing any differential and systematic object category-specific oscillatory modulation to cortical areas, despite the poor spatial resolution of EEG and the impossibility of localizing the cortical generators of scalp electrical activity with perfect confidence.

Another experimental design choice we made in the Pilot Experiment was to crop all target stimulus images and place them entirely within a uniform oval presented at the center of the image, against a black background. This design choice was intended to address the possibility that in Experiment 1, spatial attention, rather than object-based attention, was driving the systematic differences in alpha topographies that we observed. In Experiment 1, face image targets and tool image target were centrally presented and did not extend to the edges of the image boundary, whereas scene image targets were full-frame. Therefore, it is possible that when cued for faces or tools, participants would engage spatial attention in anticipation of the target and selectively process visual input from the center of the presentation area, where the cued stimulus would appear, whereas when instructed to anticipate scenes, participants' spatial attention would be oriented to a larger and more diffuse spatial region, spreading uniformly over the entire target presentation area. Because spatial attention is known to systematically modulate alpha band activity (Worden et al., 2000),

Finally, in Experiment 1, all face images were white-ethnicity, and so in Experiment 2 we drew our face stimuli from a face image database that was more comprehensive in its inclusion of multi-ethnic faces, because there was no experimental design consideration that required us to use face image stimuli representing only a single ethnic group.

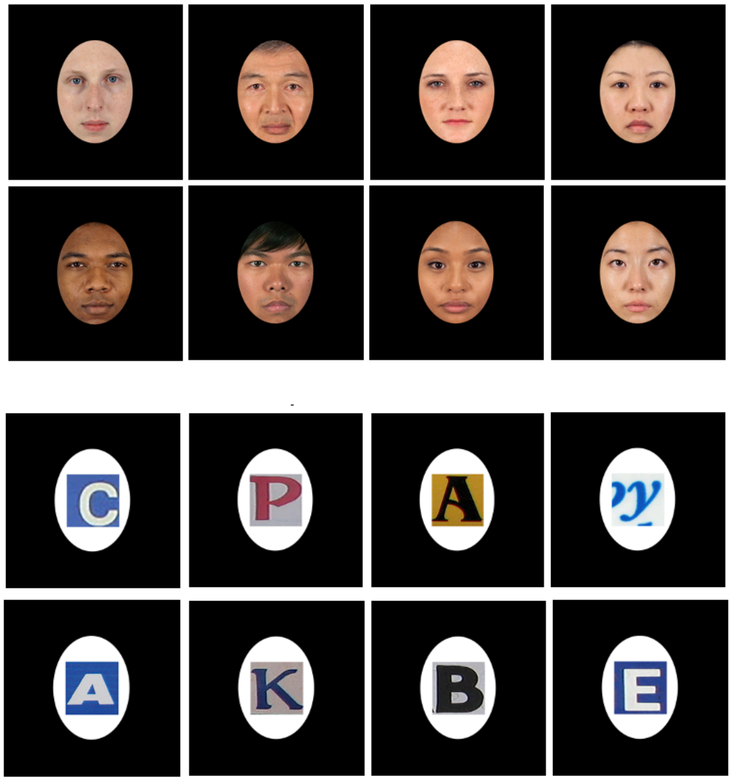
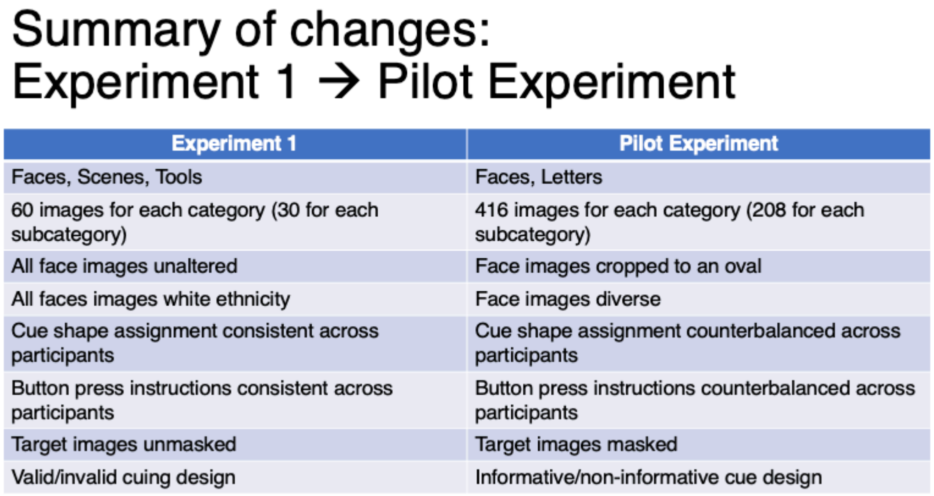


Figure 2. Examples of the face and letter target images used in the Pilot Experiment.

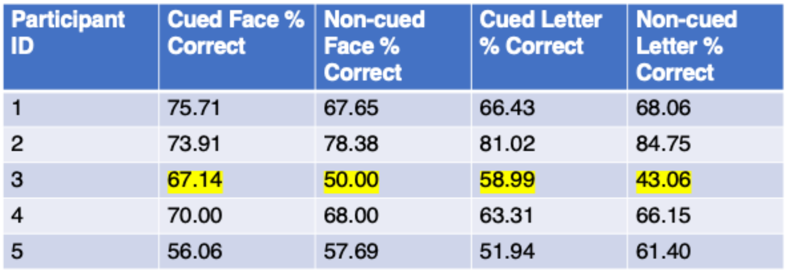


Summary of changes:
Experiment 1 → Pilot Experiment

Experiment 1	Pilot Experiment
Faces, Scenes, Tools	Faces, Letters
60 images for each category (30 for each subcategory)	416 images for each category (208 for each subcategory)
All face images unaltered	Face images cropped to an oval
All faces images white ethnicity	Face images diverse
Cue shape assignment consistent across participants	Cue shape assignment counterbalanced across participants
Button press instructions consistent across participants	Button press instructions counterbalanced across participants
Target images unmasked	Target images masked
Valid/invalid cuing design	Informative/non-informative cue design

Figure 3. Table summarizing changes in design from Experiment 1 to the Pilot Experiment.

We recorded behavioral data from five participants in order to test whether the design of this Pilot Experiment was successfully invoking object-based attention. All participants were healthy undergraduate volunteers from the University of California, Davis, had normal or corrected-to-normal vision, gave informed consent, and received course credit or monetary compensation for their time. We set an a priori stopping point of collecting data from five participants in order to assess whether the design of this experiment was achieving our goals. We found that of the five participants, only one exhibited behavioral results in the valid and invalid trials that differed in the predicted direction, with invalid trial performance indistinguishable from chance. On the basis of these preliminary behavioral results, we decided to redesign the experiment before collecting a full dataset with EEG. The design of subsequent experiments (Experiment 2 and Experiment 3) did not follow directly from the design of this Pilot Experiment, although in the case of Experiment 3, some of the same design considerations came into effect, as will be discussed below.



Participant ID	Cued Face % Correct	Non-cued Face % Correct	Cued Letter % Correct	Non-cued Letter % Correct
1	75.71	67.65	66.43	68.06
2	73.91	78.38	81.02	84.75
3	67.14	50.00	58.99	43.06
4	70.00	68.00	63.31	66.15
5	56.06	57.69	51.94	61.40

Table 1. Behavioral accuracy results from all five participants of the Pilot Experiment. Only one participant exhibited a pattern of results that accorded with our expectations for how object-based attention was operationalized in our design.

Experiment 2

In Experiments 2 and 3, we tested two alternative interpretations of our results from Experiment 1. However, these alternative interpretations were not exactly the same ones as we sought to address in the Pilot Experiment: After inspecting the results of the Pilot Experiment, we pivoted our strategy for conducting control experiment that would support Experiment 1, partially because the design of the Pilot Experiment did not elicit the kind of object-based attention behavioral results that we sought, and partially because after collecting the Pilot data, we realized that there were other potential confounds and considerations that we should methodically inspect and exclude from our interpretation of Experiment 1's results.

The Pilot Experiment was primarily designed to control for the possibility that in Experiment 1, selective anticipatory attention was not engaged at all. This possibility was addressed in the design of Experiment 3, which will be discussed below. Experiment 2 was designed to address a confound in the design of Experiment 1 that we realized would be extremely important to rule out as the primary driver of our decoding results. Namely, in Experiment 2, we tested whether decoding accuracy in the preparatory period between the cue onset and the target onset found in Experiment 1 might have been based on differences in the sensory processes evoked in the visual system by the different cue stimuli. In Experiment 1, the physical stimulus properties of the cues for the three different object attention conditions differed from one another (triangle vs. square vs. circle). It is reasonable to expect that the different physical stimulus properties of these shapes would lead to different patterns of activity in the visual system, and therefore contribute to the decoding results. In Experiment 2, we sought to investigate the temporal extent of this decoding confound.

Participants

All participants were healthy undergraduate volunteers from the University of California, Davis, had normal or corrected-to-normal vision, gave informed consent, and received course credit or monetary compensation for their time. EEG data were recorded from 29 undergraduates; datasets from four participants were rejected on the basis of irreconcilable noise in the data or subject non-compliance, yielding a final dataset from 25 participants (9 males and 16 females) that was used for further decoding analysis.

Experimental design

The study used a within-subjects design. We investigated the distributions of EEG alpha power at the scalp during the post-cue period when the three object categories were not attended in advance. Details of the task and the statistical analyses are presented in the following.

Statistical analysis

Behavioral response data were analyzed with a gamma-distributed generalized linear mixed model (Lo and Andrews, 2015) with random effect of subject and fixed effects of object category and cue validity to quantitatively assess the effect of cue validity on RT.

Differences in EEG alpha power scalp topographies as a function of cue condition were statistically analyzed using a SVM decoding approach and a non-parametric cluster-based permutation test and Monte Carlo simulation. A cluster-based statistical test was used in order to control for multiple comparisons issues that arise when t-tests are performed at all time points over the epoch (Bae and Luck, 2018). The details of the statistical test for EEG alpha power are described in the following.

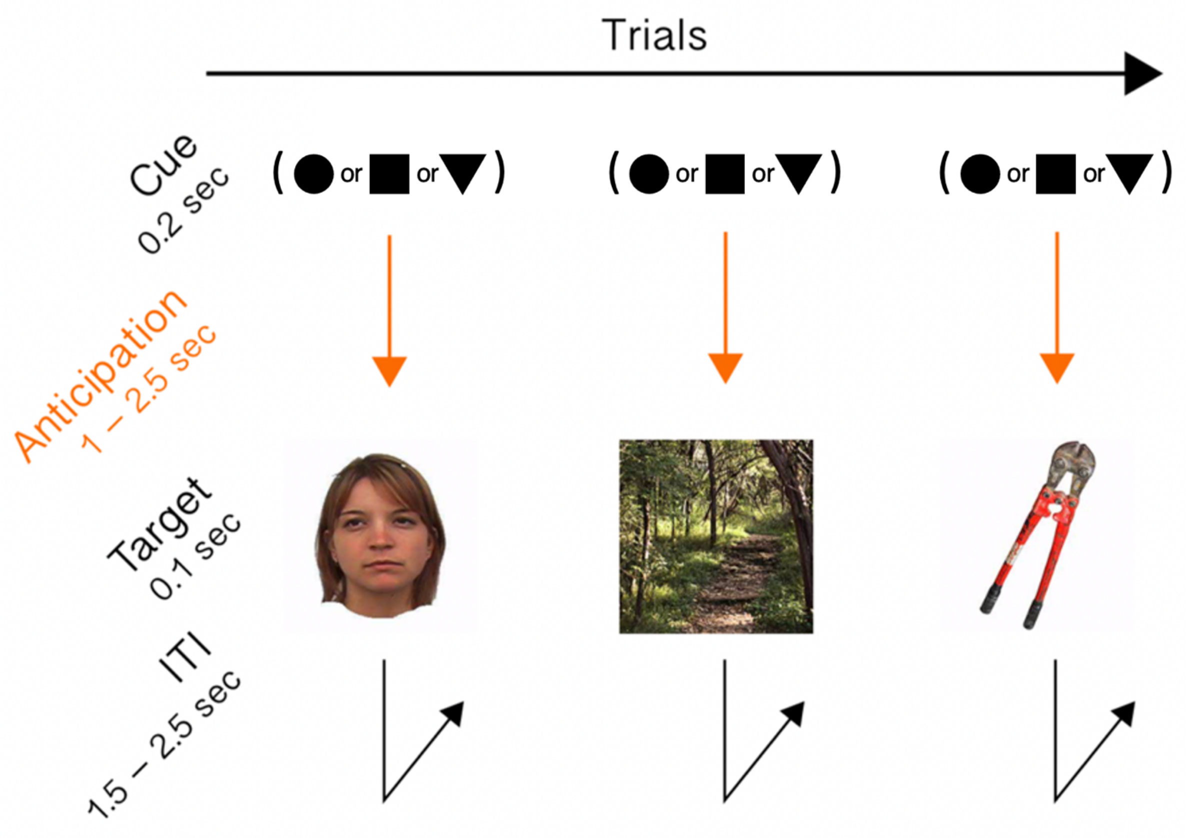


Figure 4. Example trial sequence for the second experiment. Each trial began with the presentation of a symbolic cue that was not predictive of the upcoming object category. Following an anticipation period (cue-to-target) varying from 1.0 to 2.5 s, a picture of an object (face, scene or tool) was presented. Participants were required to make a rapid-accurate discrimination of aspects of the target pictures.

Procedure

The recording and analysis protocols were identical to those of Experiment 1. Given that the purpose of this experiment was to test whether decoding accuracy in the preparatory period between the cue onset and the target onset might have been based on differences in the sensory processes evoked in the visual system by the different cue stimuli, we modified Experiment 1 by making the cues non-predictive of the upcoming target category. In keeping with this modification, we instructed participants that the cue shape was not informative, and the cue presentation was simply to alert them that the target stimulus would soon appear. Participants were not explicitly instructed to ignore the cue shape. While the time course of differences in sensory responses in scalp EEG filtered to alpha band frequencies is difficult to gauge, on the basis of the previous literature (Bae and Luck, 2018), we predicted that even for alpha, any differentiable stimulus-evoked sensory activity would be restricted to a window of time within 200 msec after the cue onset. Each participant completed 10 blocks of the experiment, with each block comprising 42 trials.

SVM decoding results

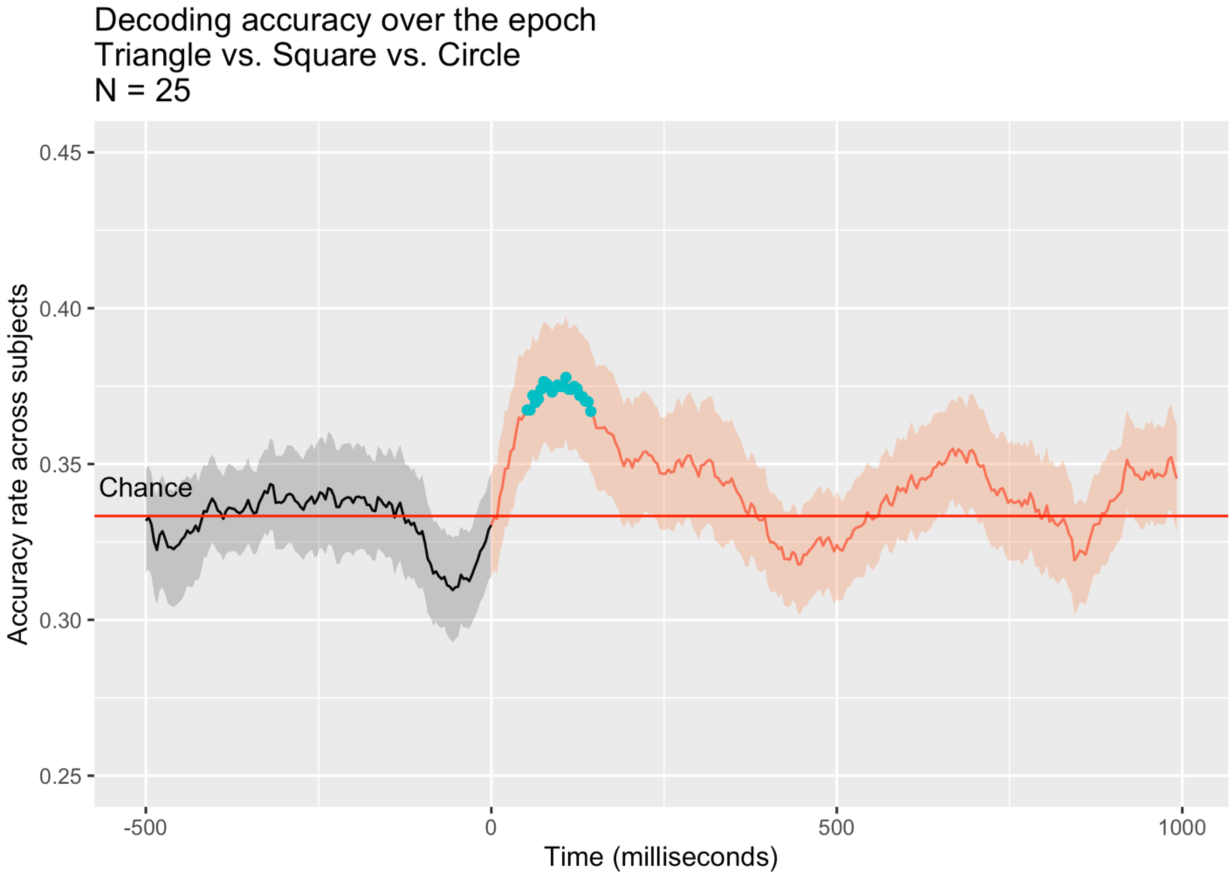


Figure 5. Alpha Band Decoding Accuracy for Experiment 2. The same SVM decoding procedure and Monte Carlo statistical procedure that was used for analyzing the data from Experiment 1 was applied to alpha band EEG from Experiment 2, revealing a cluster of statistically significant time points close to the onset of the cue, but not later in the preparatory period.

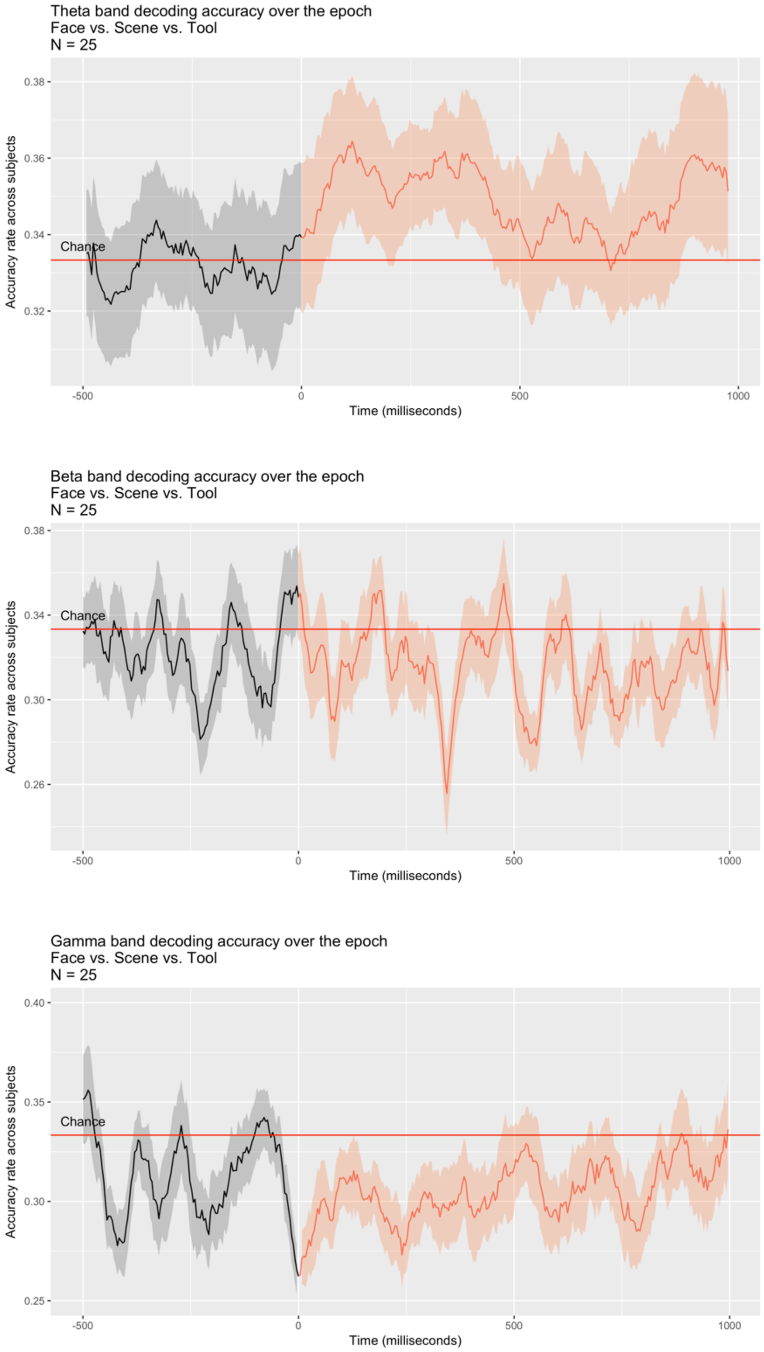


Figure 6. Theta, Beta, and Gamma Band Decoding Accuracy for Experiment 2. The same SVM decoding procedure and Monte Carlo statistical procedure that was used for analyzing the data from Experiment 1 was applied to the alternative frequency band EEG from Experiment 2, revealing no clusters of statistically significant time points.

We observed a statistically significant cluster of above-chance decoding accuracy timepoints in the cue presentation window only. No further clusters of significantly above-chance decoding occurred anywhere from 200 msec to 1000 msec (Figure 5). We also observed that there were no statistically significant clusters of above-chance decoding accuracy timepoints for theta band, beta band, and gamma band filtered data (Figure 6).

The results of this control experiment argue against the possibility that the late-period alpha band decoding we observed in our original experiment was simply a result of differential bottom-up sensory processes across the three cue conditions. Because the paradigm for Experiment 2 was identical to the paradigm of Experiment 1 in every respect other than the cue validity, and because we ran the same SVM decoding pipeline on the alpha band EEG data from Experiment 2 as we did in Experiment 1, we could directly assess whether the pattern of decoding results we obtained from the original experiment was attributable to bottom-up sensory processes.

We collected data from more participants for Experiment 2 than we did for our original experiment so that we could have more power in assessing the magnitude and the temporal extent of the decoding that could be achieved purely on the basis of stimulus-evoked activity. Our results support the idea that the long-latency above-chance decoding in Experiment 1 is not attributable to purely sensory activity driven by physical stimulus differences, because we found that in Experiment 2, statistically significant above-chance decoding occurred only in a cluster of time points at short post-cue latency (< 200 msec after cue onset; Figure 5).

Experiment 3

In Experiments 2 and 3, we tested two alternative interpretations of our results from Experiment 1. In Experiment 3, we investigated whether differences in alpha topography across object attention conditions in Experiment 1 may have been the result of different task sets across the three object attention conditions, rather than reflecting object-based attention mechanisms in visual cortex.

Participants

All participants were healthy undergraduate volunteers from the University of California, Davis, had normal or corrected-to-normal vision, gave informed consent, and received course credit or monetary compensation for their time. EEG data were recorded from 23 healthy undergraduate volunteers (5 males and 7 females). Datasets from three participants were rejected on the basis of irreconcilable noise in the EEG data or subject noncompliance, yielding a final dataset of EEG data from 20 participants (9 males and 11 females) that was used for further decoding analysis.

Experimental design

The study used a within-subjects design. We investigated the distributions of EEG alpha power at the scalp, as a function of attended object category, in an anticipatory cued attention task with three categories of objects (faces, scenes and tools). Details of the cued object-based attention task are presented in the following.

The recording and analysis protocols were identical to those of Experiment 1. The purpose of this experiment was to investigate whether differences in alpha topography across object attention conditions in Experiment 1 may have been the result of different task sets across the three object attention conditions, rather than reflecting object-based attention mechanisms in visual cortex. Specifically, in the attend-face condition of Experiment 1, participants were instructed to discriminate whether the presented face was male or female, and indicate their choice using a button box with two buttons under the index finger and middle finger. In the attend-scene condition, the task was to discriminate urban from natural scenes using the same two buttons, and in the attend-tool condition, the task was to discriminate powered from non-powered tools using the same two buttons. Because the categories being discriminated were different across the different cue conditions (male/female, urban/natural, power tool/hand tool), it is possible that participants were preparing different task sets across the different cue conditions during the preparatory period. After being presented with a triangle cue, for example, a participant would need to cognitively map their index finger response to the identification of a male face and their middle finger response to the identification of a female face, whereas this mapping would be different if the participant were presented with a square cue. These different task sets and mappings from visual cortex to motor response preparation could possibly have been driving the different alpha scalp topographies over the preparatory period.

This explanation is not mutually exclusive of our interpretation that alpha scalp topographies reflect differential preparatory attentional biasing in object category-selective visual areas, but given the design of Experiment 1, there is no way to know whether one, the other or both are reflected in the differing alpha patterns. Therefore, we conducted an experiment that equated the task across all object attention conditions, in order to eliminate any task set differences that were present in the original experiment. Based our model that alpha is a mechanism for selective attention to objects in visual cortex, in this new design we should still observe different patterns of alpha for preparatory attention to object categories, which should be revealed in successful decoding late in the cue-to-target period.

Apparatus and stimuli

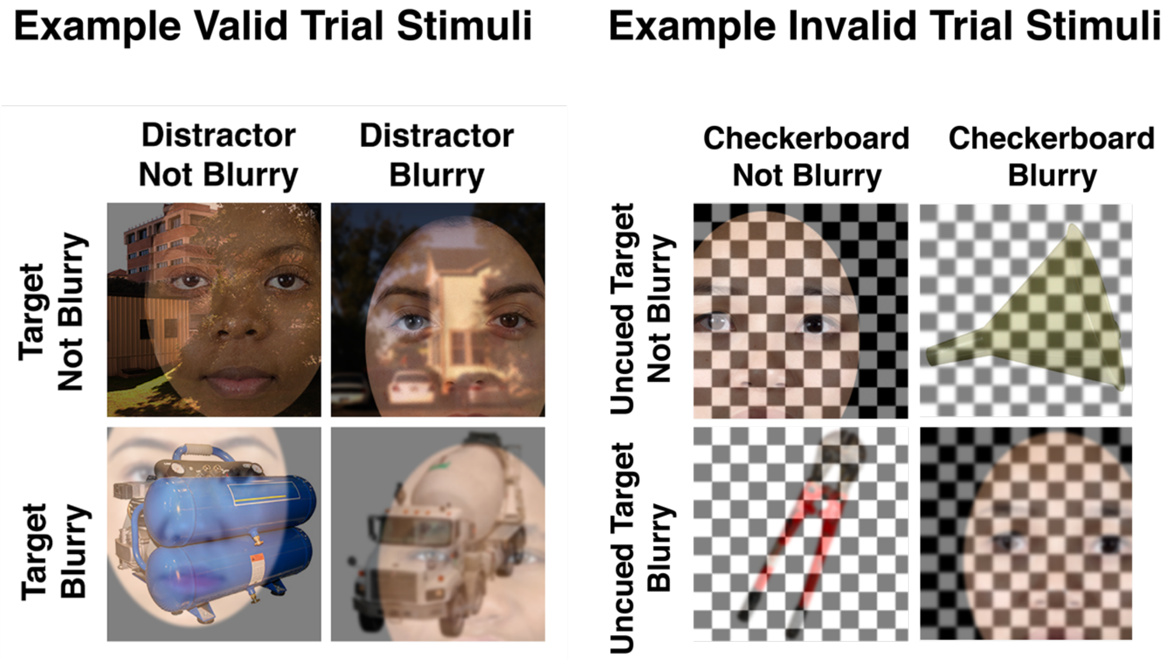


Figure 7. Example stimulus images in the attention task for Experiment 3. In the set of example valid trial stimuli shown, Face is the target object category to be identified as in-focus or blurry, and the overlaid tool or scene images are the distractor images. For each stimulus image, both the target and distractor can be blurry or in-focus, independently of each other. Example invalid trial stimuli are also provided to illustrate that both the uncued target

image and the overlaid checkerboard can be blurry or not-blurry, independently of one another. In the invalid trial condition, participants were still trained to respond to the uncued target image with the same blurry/not-blurry distinction, using the same response buttons as for valid trials.

The general structure of the paradigm for Experiment 3 followed the paradigm of the Experiment 1. On each trial, a cue shape appeared, indicating the object category to attend. Cue shapes were identical to those in Experiment 1. As before, a preparatory period followed the cue, and then a stimulus image appeared. An ITI separated the stimulus image and the onset of the next trial. Behavioral responses were collected during this ITI. SOA and ITI ranges were kept the same as in Experiment 1.

The behavioral task for this experiment was to determine, on each trial, whether the briefly presented target image belonging to the cued object category (faces, scenes, or tools) was in-focus or blurry. Unlike Experiments 1 and 2, the stimuli to be discriminated were composites of an image belonging to the target category superimposed with an image belonging to a non-cued, distractor category. Crucially, both the target image and the distractor image in the blend could be in-focus or blurry independently of each other, therefore, the task could not be performed solely on the basis of attending to and responding to the presence or absence of blur (Figure 7).

Twenty percent of trials were invalidly cued, allowing us to assess the effect of cue validity on behavioral performance. For the invalid trials, the stimulus image was a composite of an image from a randomly chosen non-cued object category, superimposed with a black and white checkerboard. The checkerboard could also be blurry or in-focus independently of the object image. Participants were instructed that whenever they encountered a trial where the blended stimulus didn't include an image belonging to the cued object category, but instead contained only one object image and a checkerboard overlay, then they had to indicate whether the non-cued object image in the stimulus was blurry or in-focus. We predicted that participants would be slower to respond on invalidly cued trials, analogously to the behavioral effect of validity observed in cued spatial attention paradigms.

The stimulus images spanned a square 5° x 5° of visual angle. To create blurred images, Gaussian blur with a standard deviation of 2 was applied to the images.
All three object categories included 40 different individual images. On each trial, random images were drawn to produce the composite stimulus image. Scene and tool images were drawn from the same image sets as those for the original experiment. However, face images were drawn from a different image set (Ma et al., 2015) because the face images used in the original experiment were not high enough resolution to yield reliably noticeable differences in blurred vs. non-blurred conditions. All face images were cropped to ovals centered on the face and placed against a white background.

Unlike scene images, which contained visual details spanning the entire 5° x 5° square, face and tool images were set against white backgrounds and so did not contain visual information up to all the image boundaries. Therefore, to eliminate the possibility that participants could use cue information to focus spatial attention instead of object-based attention to perform the blurry/in-focus discrimination, on any trial where a face or tool image was included in the composite stimulus, the position of that face or tool image was randomly jittered from the center.

Procedure

Participants were instructed to respond as quickly as they could to the target stimulus, making it vital that the participants engaged preparatory attention toward the cued object category during the preparatory period. All participants were trained with at least 42 trials of the task, and were able to achieve at least 60% response accuracy before performing it under EEG data collection; to achieve this, stimulus duration was adjusted on an individual participant basis during the initial training phase. Experiment 3 was conducted in the same laboratory environment as the original experiment, and environmental setup variables were equated to those of the original experiment. Each participant completed 15 blocks of the experiment, with each block comprising 42 trials, which represented, on average, 210 more trials per subject than Experiment 1.

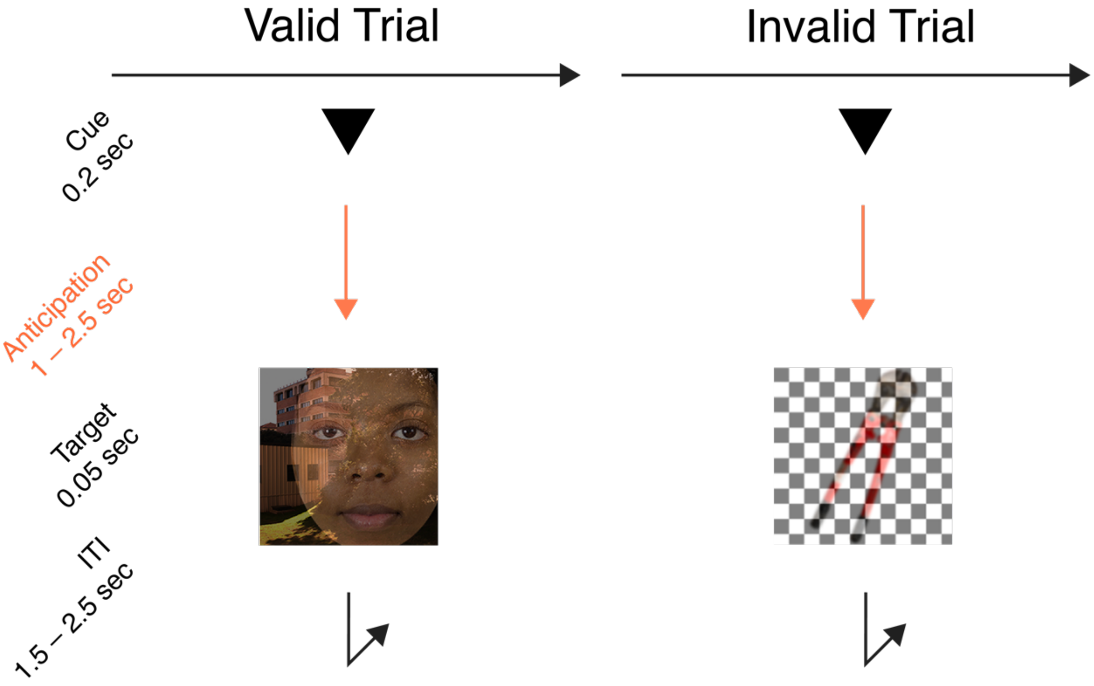


Figure 8. Example trial sequence in Experiment 3. Each trial began with the presentation of a symbolic cue that was predictive of the upcoming object category (75%). Following an anticipation period (cue-to-target) varying from

1.0 to 2.5 s, a composite stimulus image was presented. Participants were required to make a rapid-accurate discrimination of whether the cued object image was blurry or not-blurry (valid trials), or whether the uncued object image was blurry or not-blurry (invalid trials).

Behavioral results

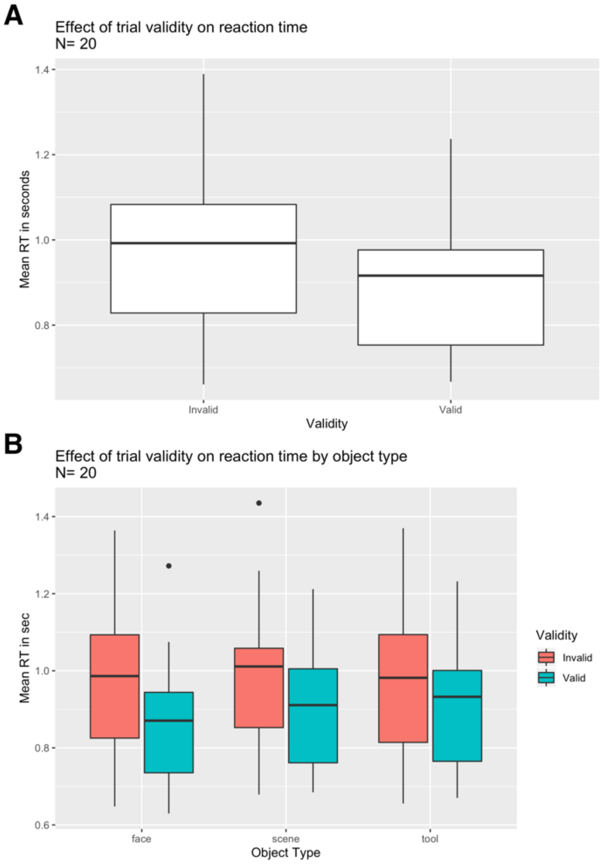


Figure 9. Behavioral Measures of Attention in Experiment 3. A. Box plots of reaction time data for invalid and valid trials, collapsed across 20 subjects, averaged across attention (object) conditions. Thick horizontal lines inside boxes represent median values. First and third quartiles are shown as lower and upper box edges. Vertical lines extend to most extreme data points excluding outliers. Dots above plots represent outliers, defined as any value greater than the third quartile plus 1.5 times the interquartile range. Subjects were significantly faster overall for cued (valid) objects than uncued (invalid) objects. **B.** Reaction times for valid and invalid trials separately for each attention condition. Subjects were significantly faster for cued (valid) objects than uncued (invalid) objects for each object category.

We observed differences in RT between valid and invalid trials, for all object categories, such that validly cued trials elicited faster responses than invalidly cued trials (Figure 9). In fitting a gamma-distributed generalized linear mixed model to the RT data, we found a significant effect of validity ($p < 0.001$).

SVM decoding results

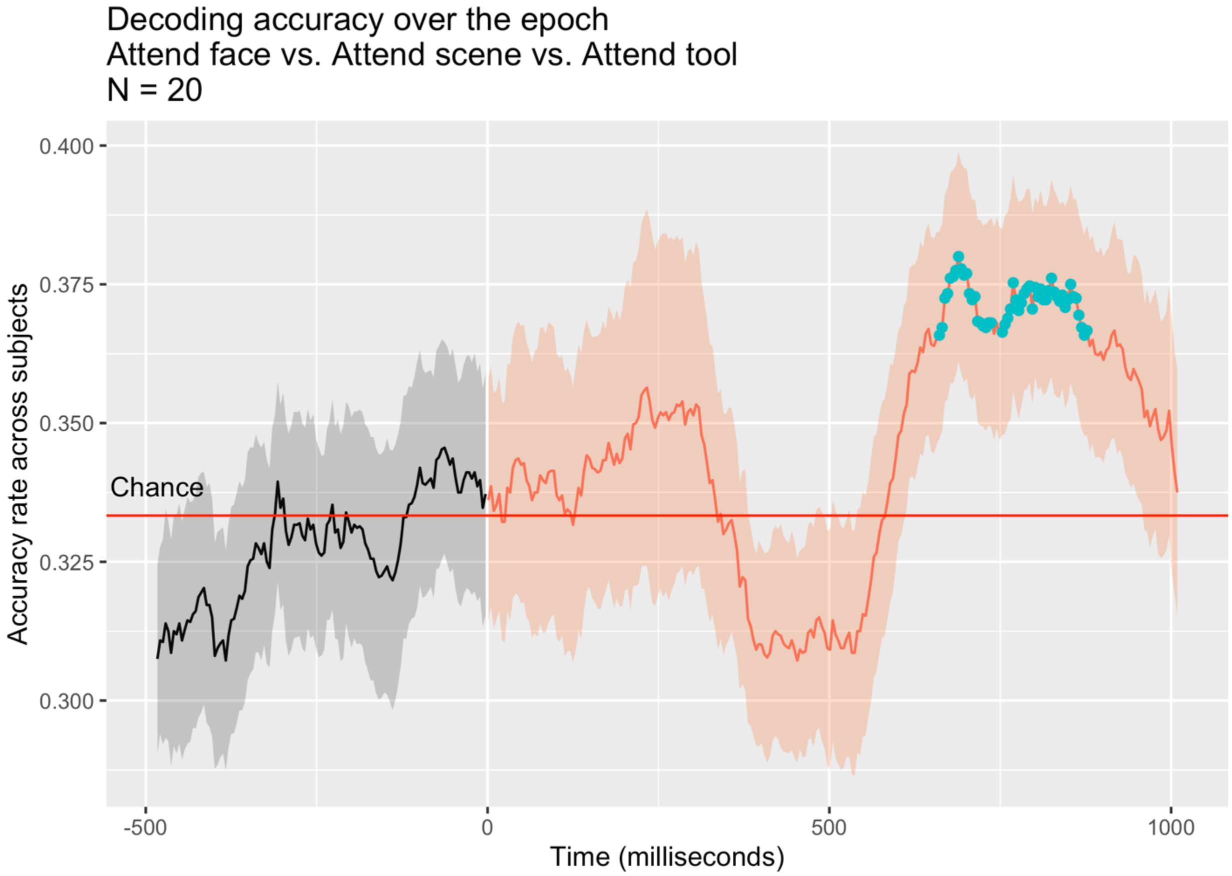


Figure 10. Alpha Band Decoding Accuracy for Experiment 3. The same support vector machine decoding procedure and Monte Carlo statistical procedure that was used for analyzing the data from Experiment 1 was applied to alpha band EEG from Experiment 3, revealing a cluster of statistically significant time points in the second half of the preparatory period.

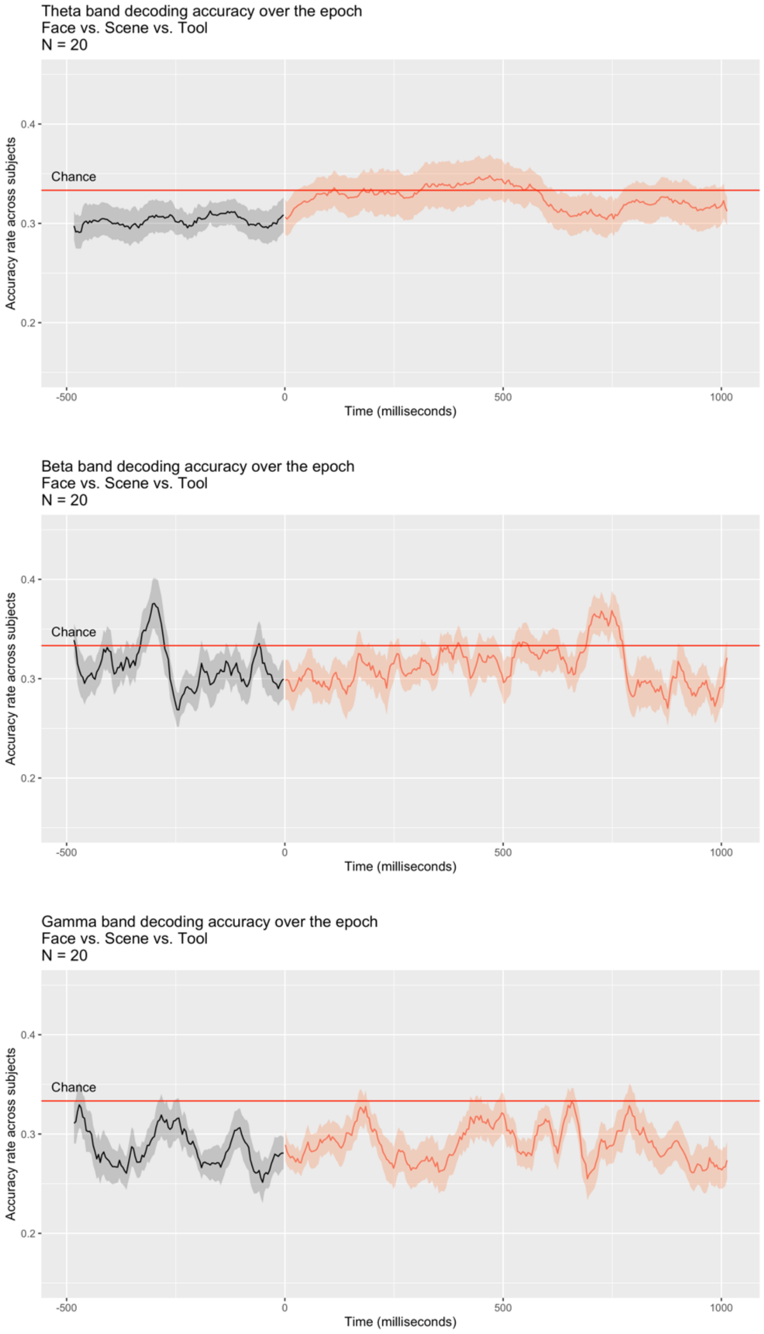


Figure 11. Theta, Beta, and Gamma Band Decoding Accuracy for Experiment 2. The same SVM decoding procedure and Monte Carlo statistical procedure that was used for analyzing the data from Experiment 1 was applied to the alternative frequency band EEG from Experiment 3, revealing no clusters of statistically significant time points in any other frequency band.

Using the same EEG analysis and SVM decoding pipeline as for Experiment 1, we found statistically significant clusters of timepoints exhibiting above-chance decoding accuracy (Figure 10). Just as in Experiment 1, these statistically significant clusters were observed in the second half of the preparatory period, more than 500 msec after the cue onset. Notably, there also appears to be a group of above-chance time points in the cue presentation window of 0 – 200 msec, in the same period where we observed statistically significant decoding in Experiment 2 that was attributable to the cue-evoked sensory activity. However, in the results of Experiment 3, like Experiment 1, decoding in this cue presentation time period (< 200 msec latency) did not reach the level of statistical significance (whereas with the higher number of participants in Experiment 2, it could be revealed).

The behavioral results of Experiment 3 suggest that participants were engaging object-based attention during the preparatory period. Participants were faster to discriminate object images as blurry or in-focus when their category was cued. Analogously to cued spatial attention paradigms, on invalidly cued trials, participants were attending to one object category during the preparatory period, but then upon stimulus presentation, reoriented their attention to be able to discriminate whether an image from an uncued object category was blurry or in-focus.

With the behavioral effect between valid and invalid trials in line with that from our original experiment, we are confident that the experimental design in Experiment 3 was engendering the same form of top-down object-based attention as was captured by Experiment 1. Therefore, in observing statistically significant above-chance decoding in the same general window of time after cue onset for Experiments 1 and 3, we interpret this finding as evidence that object-based attention, and not task set or motor response preparation differences, is driving the longer-latency decoding result prior to onset of the targets.

The decoding results for the theta, beta, and gamma bands did not reveal any statistically significant time points, suggesting that activity in these frequency bands is not systematically modulated by the conditions of the experiment (Figure 11).

SVM Decoding of Random Data

From the alpha power SVM decoding figures, it can be observed that there are periods of time where decoding appears to be considerably below theoretical chance (33%). For example, this dip below chance can be seen in the decoding figures for Experiment 1 and Experiment 3 during the baseline period before the onset of the instructional cue. Our interpretation of these below-chance decoding accuracy periods is that they arise entirely by chance. Indeed, previous SVM decoding studies have asserted that there is no logical way to interpret below-chance decoding accuracy (Bae and Luck, 2018).

Equally puzzling are the periods of above-chance decoding accuracy in the baseline period before the onset of the instructional cue. Before the cue, there is not any information available to the participants about the identity of the upcoming cue, because the order of trials is completely randomized. Therefore, decoding accuracy during this baseline period should be equal to chance in theory, because there should be no systematic information about the object condition of each trial available to the brain during this time. Because of the apparent discrepancies from theoretical decoding behavior observed in our various decoding results, we wanted to carefully assess whether there was any aspect of the decoding pipeline that might be biasing the analysis and distorting our decoding accuracy results.

The most straightforward way to check for bias in the SVM pipeline is to replace the actual EEG data with random noise and run this pure-noise simulated data through the same decoding pipeline and cluster-based permutation test for statistical significance as was used for our experimental data. If when operating over random data decoding accuracy values are centered on theoretical chance, and no clusters of statistically significant time points emerge from the statistical test, then it can be assumed that there is no systematic bias in the analysis, because random data by definition and in implementation contains no information about the condition labels that are arbitrarily assigned to it at any point in the decoding epoch. A further useful outcome would be that testing the decoding method on random noise would allow us to gauge the extent of deviation from theoretical chance that might be expected under a null hypothesis that there was no meaningful experimental condition information in the data that was submitted to the analysis.

We therefore re-ran the exact SVM decoding analysis and statistical analysis used in Experiment 1, Experiment 2, and Experiment 3, but for each participant replaced real EEG data with randomly generated data. The result of this control analysis is provided below in Figure 12.

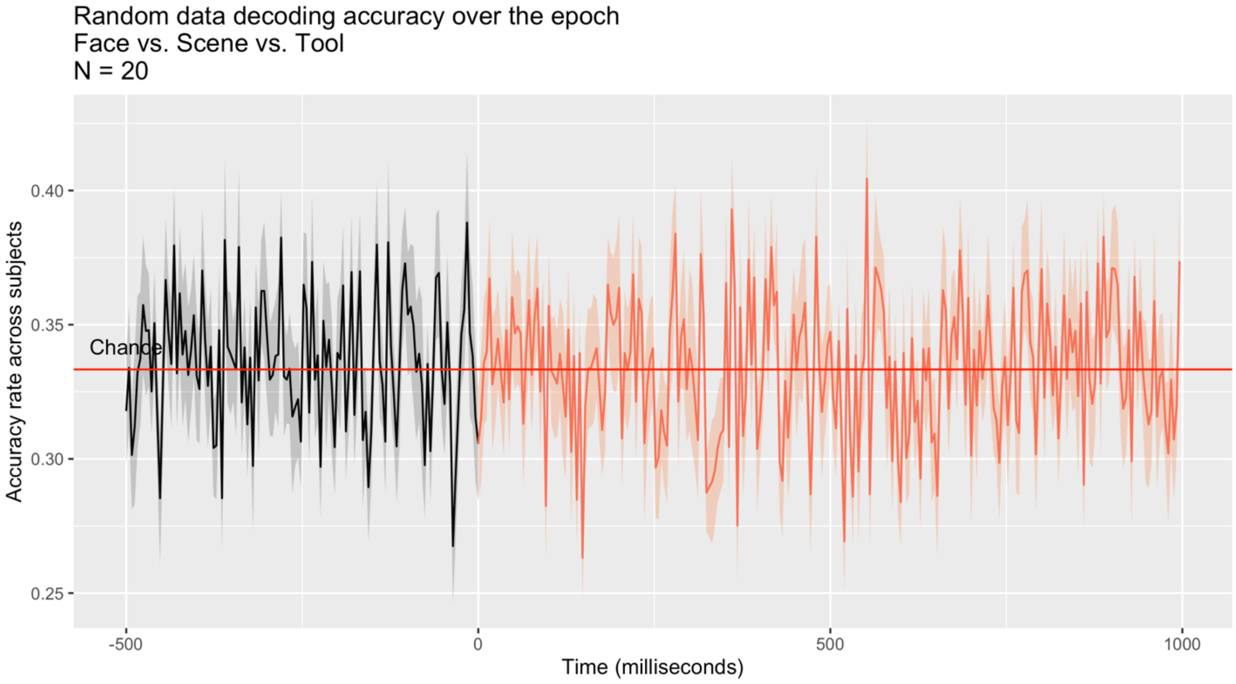


Figure 12. Decoding accuracy time series for SVM analysis of random noise. We ran the SVM decoding pipeline again, replacing EEG data for each subject with random noise, to test for bias in the analysis. Because the resulting decoding accuracy is centered on theoretical chance (33%), we can rule out the possibility that there is bias in the analysis pipeline, strengthening our interpretation of the decoding results from Experiments 1–3.

We found that the decoding accuracy time series for this random data centered on theoretical chance, validating our expectation that there was no systematic bias in the SVM decoding pipeline or in the cluster-based permutation test for statistical significance, and strengthening our confidence in the decoding results for Experiments 1, 2, and 3. Furthermore, the random decoding accuracy tended to oscillate between peaks about 5% above and below chance. From this behavior we have a shorthand measure of how much decoding accuracy might be expected to deflect from the theoretical chance level of exactly 33% under the null hypothesis.

Experiment 3 Behavioral Control Experiment

In Experiment 3, object-based attention was operationalized with a cueing paradigm, in which on each trial the appearance of one of three possible object categories (faces, scenes, or tools) was indicated ahead of time with an 80% predictive cue. For the 80% of trials that were validly cued, an image from the cued object category appeared superimposed with an image from an uncued object category. For the 20% of trials that were invalidly cued, an image from an uncued object category appeared superimposed with a checkerboard pattern instead of another object image (Figure 7; Figure 8). The invalid trial stimuli were designed such that only one object image was present, to preclude any ambiguity about what object image was the intended target of the discrimination task in the absence of an image from the cued object category. The checkerboard pattern was displayed overlaid with the invalidly cued object image to provide distracting visual information comparable to that from an overlaid uncued object image in the valid trials.

This paradigm was designed so that the effect of cued, anticipatory object-based attention would manifest as a difference in reaction time between valid and invalid trials. We predicted that invalid trials would display longer reaction times than valid trials, because the beneficial effects of anticipatory attention would only apply when the anticipated object category was present in the target stimulus.

However, the systematic difference in visual stimulus characteristics between the valid and invalid trials, with valid trial stimuli composed of two object images and invalid trial stimuli composed of a checkerboard pattern and one object image, could lead to the question of whether the observed behavioral decrement on the invalid trials was truly the result of anticipatory object-based attention to a different object category, or if instead the invalid trial design was simply more difficult to visually parse. It is reasonable to wonder whether the checkerboard pattern overlaid on the uncued object image targets made discrimination of those targets more difficult than discrimination of an object image in superimposition with another object image.

We conducted an experiment to address this concern and to support our interpretation that the difference in reaction times between valid and invalid trials reflects the behavioral benefit of object-based attention. We collected behavioral data from ten participants (4 females, 5 males, 1 non-binary). All participants were healthy undergraduate and graduate volunteers from the University of California, Davis community, had normal or corrected-to-normal vision, gave informed consent, and received monetary compensation for their time.

The design and procedure of this experiment were identical with those of Experiment 3, with one crucial modification: Twenty percent of the trials were validly cued, but the stimulus was composed of an object image from the cued object category superimposed with the same checkerboard pattern present in invalid trial stimuli. We designate these new trials as valid checkerboard trials. Thus 60% of trials were the same validly cued trials from Experiment 3, 20% of trials were the same invalidly cued trials from Experiment 3, and 20% of trials belonged to the new valid checkerboard condition.

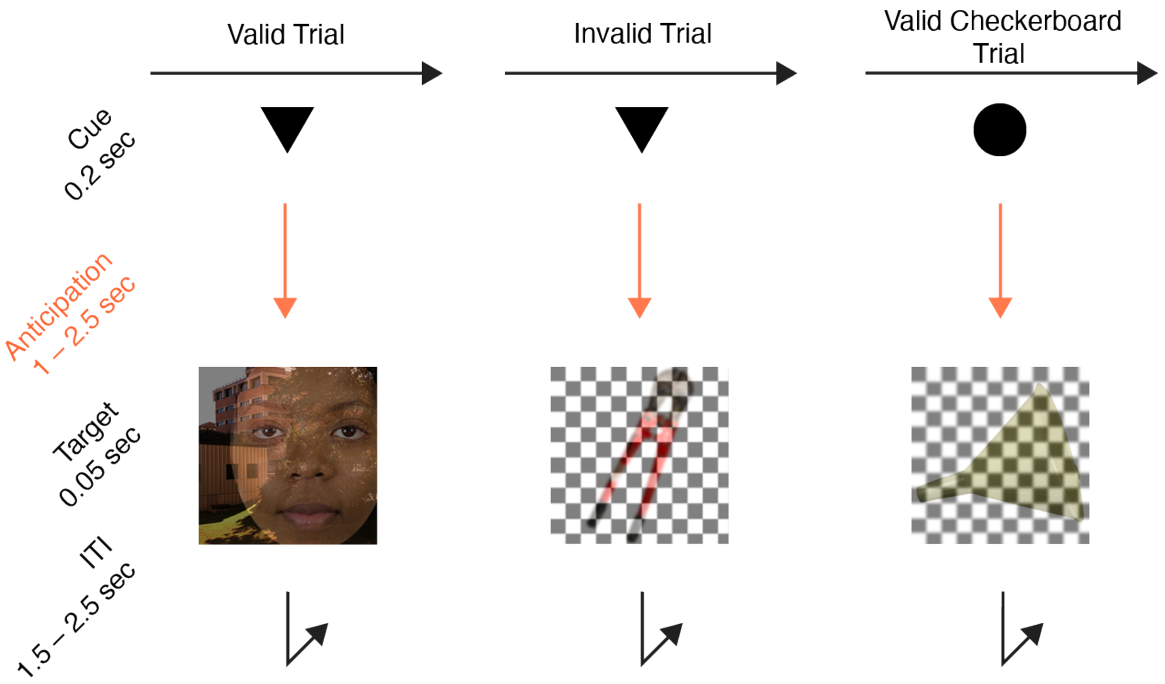


Figure 13. Example trial sequence in behavioral control experiment. Each trial began with the presentation of a symbolic cue that was predictive of the upcoming object category (75%). Following an anticipation period (cue-to-target) varying from 1.0 to 2.5 s, a composite stimulus image was presented. Participants were required to make a rapid-accurate discrimination of whether the cued object image was blurry or not-blurry (valid and valid checkerboard trials), or whether the uncued object image was blurry or not-blurry (invalid trials).

The introduction of valid checkerboard trials in this experiment allows us to examine whether the Invalid trials in Experiment 3 elicited longer reaction times than valid trials simply because of the presence of the checkerboard in the stimuli. If the presence of the checkerboard pattern is driving the longer reaction times in the invalid trials of Experiment 3, then in the present experiment, the valid checkerboard condition will also elicit longer reaction times than the valid condition.

The results of this experiment are displayed in Figure 14 and Figure 15 below. Invalid trials display longer reaction times than both valid and valid checkerboard trials (Figure 14). This result was consistent regardless of which of the three object categories was the target on a given trial (Figure 15). Furthermore, the valid checkerboard reaction times were not significantly faster than valid reaction times. Welch's two-sample *t*-test between: 1. Invalid and valid RTs indicates a significant difference ($p < 0.01$); 2. Invalid and valid checkerboard RTs indicates a significant difference ($p < 0.01$); 3. Valid and valid checkerboard RTs indicates no significant difference ($p = 0.4$). Together, these results indicate that the checkerboard pattern by itself did not hamper the behavioral performance of the participants in Experiment 3, and thus the relatively faster reaction times observed in valid trials can be attributed to the beneficial effects of anticipatory object-based attention.

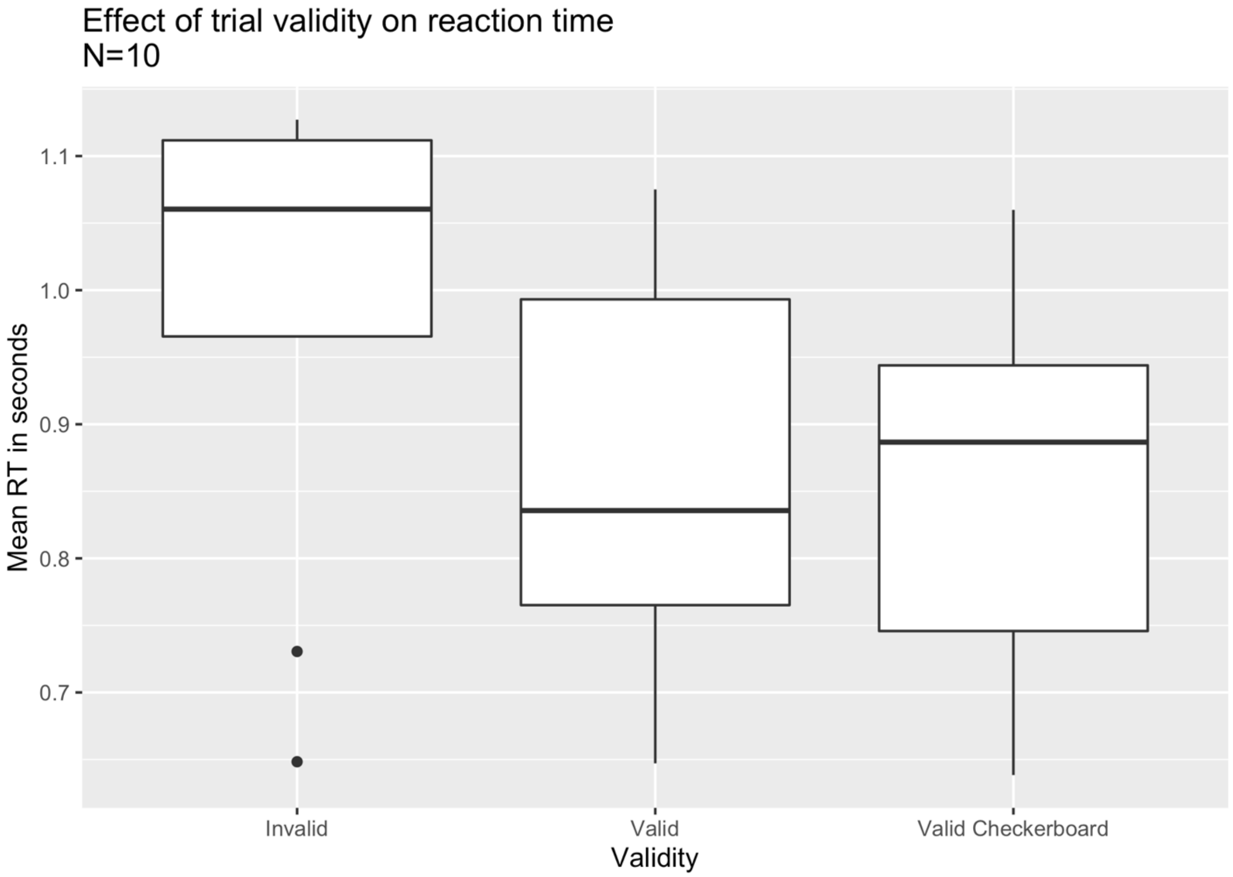


Figure 14. Box plots of reaction time data for invalid, valid, and valid checkerboard trials, collapsed across 10 subjects, averaged across attention (object) conditions. Thick horizontal lines inside boxes represent median values.

First and third quartiles are shown as lower and upper box edges. Vertical lines extend to most extreme data points excluding outliers. Dots above or below plots represent outliers, defined as any value greater than the third quartile plus 1.5 times the interquartile range. Subjects were significantly faster overall for cued (valid) objects than uncued (invalid) objects.

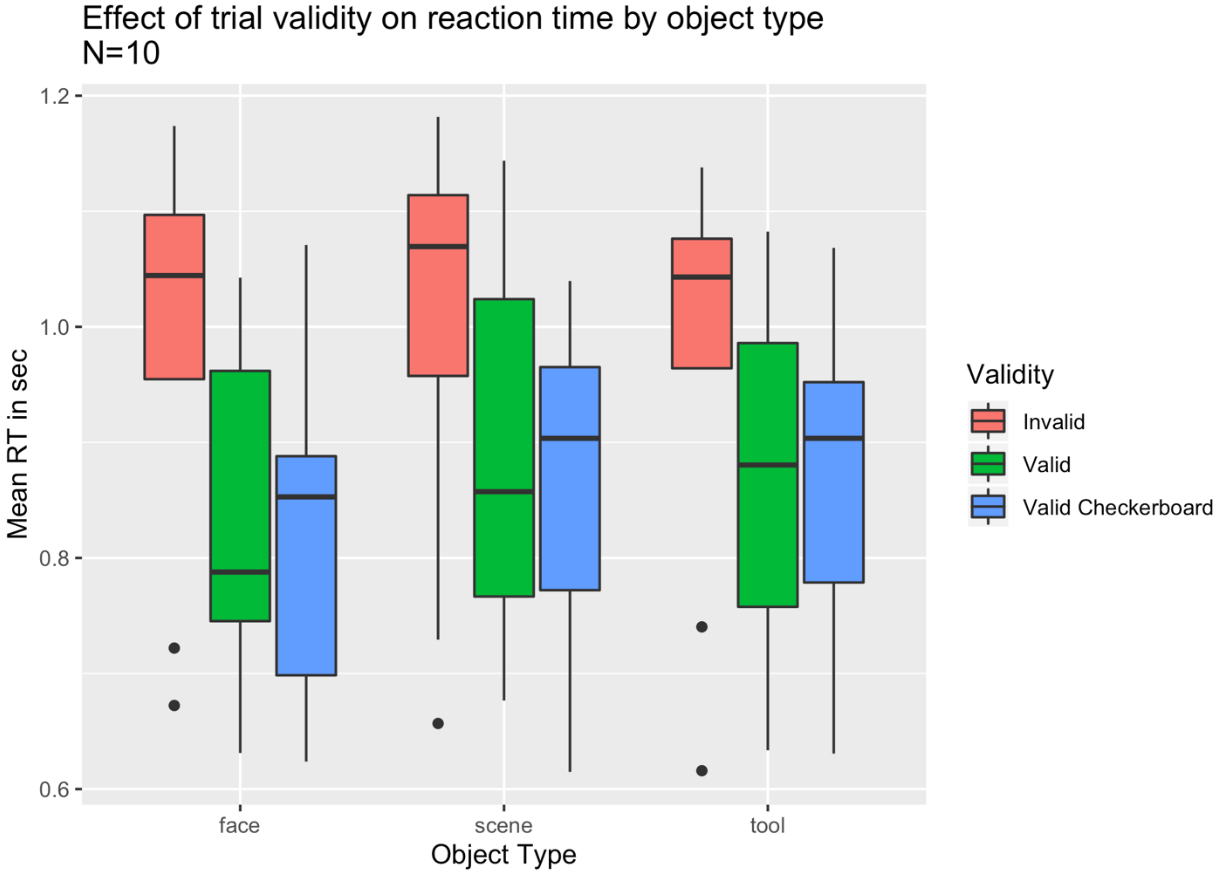


Figure 15. Box plots of reaction time data for invalid, valid, and valid checkerboard trials, collapsed across 10 subjects, displayed separately for each object condition. Thick horizontal lines inside boxes represent median values. First and third quartiles are shown as lower and upper box edges. Vertical lines extend to most extreme data points excluding outliers. Dots above or below plots represent outliers, defined as any value greater than the third quartile plus 1.5 times the interquartile range. Subjects were significantly faster overall for cued (valid) objects than uncued (invalid) objects, for each object category.

Chapter 4: ERP Decoding Analyses

Experiment 1 was designed specifically to test the hypothesis that object-based attention modulates alpha band oscillatory neural activity throughout the visual system, and that these systematic patterns of alpha band modulation are visible at the level of the scalp with EEG recording. Therefore, the primary method utilized in the analysis of the data from Experiment 1 was SVM decoding of alpha power topography, to measure the extent to which alpha patterns are reliably distinct as a function of the object category being selectively attended. Experiment 2 and Experiment 3 were designed to rule out several alternative interpretations of the results from Experiment 1, and also both utilized SVM decoding of EEG alpha power topography as their primary analysis method.

The present chapter will describe the execution and results of ERP decoding analyses that we performed after our primary scientific question had been answered. The theoretical interpretation of the results of these ERP decoding analyses will be elaborated in the following General Discussion chapter.

Experiment 1

ERP decoding of the cue period

The first ERP decoding analysis that we conducted was an SVM decoding of the object attention cue condition over phase-locked ERP voltage, instead of alpha power data. The data used for the ERP decoding was restricted to signals below 6 Hz in frequency, to minimize the overlap in information with the alpha band data that was subjected to SVM decoding in our primary analyses. This rationale was adopted from the SVM decoding analysis of Bae and Luck (Bae and Luck, 2018), in which the filtering was performed to avoid contamination from alpha band activity, which may appear as sustained ERPs under some conditions (Mazaheri and Jensen, 2008; Van Dijk et al., 2010). In order to restrict the ERP data signals to the 0 – 6 Hz frequency range, we performed a band pass filtering procedure over the preprocessed EEG data, using the *eegfilt()* function from the EEGLAB Matlab library.

A further diversion from the data processing pipeline of our primary alpha band decoding analysis was that we did not subject the low pass filtered EEG data to a Hilbert transform. In our primary analysis, the Hilbert transform processing step, performed after filtering the preprocessed EEG data to the alpha band frequency range, had the effect of providing an estimate of instantaneous alpha power at each time point in the data. The Hilbert transform effectively computes the envelope of the waveform over which it operates. In the ERP decoding analysis of the ERP's sensitivity to the cued attention conditions, we had no a priori interest in the distribution of power in the 0 – 6 Hz frequency band at each time point in the epoch, but instead were interested in the raw ERP waveforms, and the extent to which the experimental conditions could be decoded from this EEG data time locked to the cue onset.

Aside from these alterations made to the decoding procedure, the rest of the analysis exactly followed the SVM decoding analysis performed in our primary analysis. Decoding was performed independently at each time point in the epoch. Trial data from each condition were averaged together to improve the signal-to-noise-ratio, after low passing the data to isolate the sustained ERP waveforms, and model training and testing were performed on the average EEG data. Model testing was scored as either correct or incorrect, depending on whether it labeled the experimental condition correctly – thus theoretical chance was set to one-third (33%). The test for statistical significance of decoding accuracy was the same cluster-based permutation test as used in the primary analysis: The test identified temporally contiguous clusters of decoding accuracy timepoints that were deemed significantly above chance, by comparison with an a priori 95% threshold on a null distribution of cluster t masses. The same Matlab *fitecoc*() function was used to perform SVM decoding according to the error-correcting output codes method of solving the multiclass categorization problem for a binary classification model.

Just as the primary analysis of SVM decoding over alpha power scalp distributions was essentially an assessment of the information pertaining to the object-based attention condition in the alpha band data, this SVM ERP analysis is an assessment of the information content in the ERP waveform. If the pattern of ERP waveforms across the scalp elicited by the instructional cue systematically varies as a function of the cue shape or the form of attention instructed by the cue, then the decoding should return statistically above-chance labeling of data not incorporated into the model's training phase.

As was the case in Experiment 1, it is likely that any above-chance decoding results observed during the very early part of the cue period, right after cue onset, will be confounded by the inescapable fact that the different cue shapes have different stimulus properties and likely elicited different visual activity just on the basis of their physical properties. Furthermore, at the cue offset time of 200 msec, it is possible that an additional ERP waveform will be elicited by the sensory response to the offset of the stimulus (Luck, 2014), and so the time period of above-chance decoding attributable to this sensory response might be expected to extend past 200 msec. As was the case for the primary analysis, this confound could not be circumvented with a counterbalancing procedure in which participants were assigned different mappings of cue shape to attentional instruction, because all SVM decoding was performed within subject.

The results of the SVM decoding analysis of ERP waveforms are presented in Figure 1 below. The shading around the bold central line reflects standard error across subjects, and the bold line represents the mean decoding accuracy across subjects. Turquoise dots along the central line denote the statistically significant time points, where decoding accuracy across subjects was significantly greater than would be expected under a null scenario in which the ERP waveform contained no information about the cue condition.

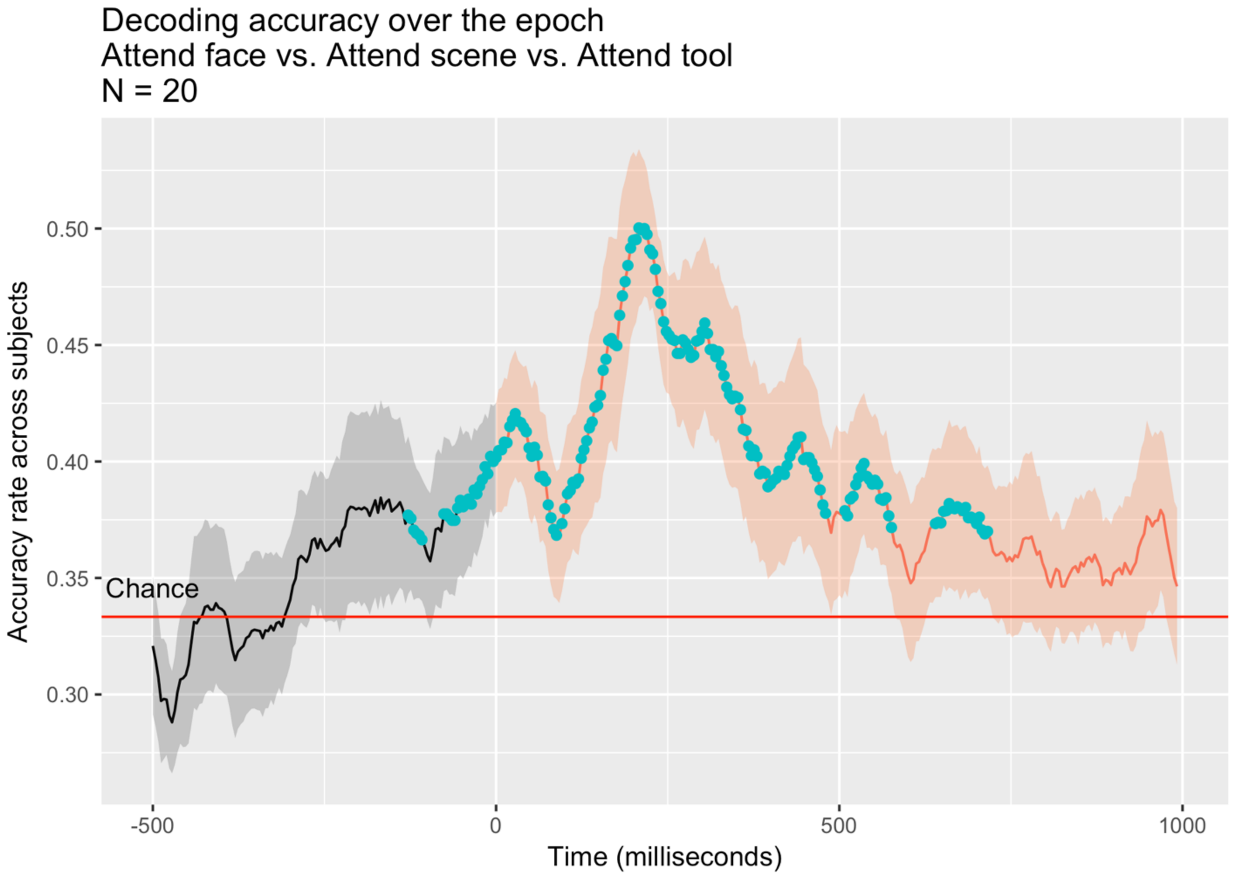


Figure 1. Experiment 1 ERP decoding accuracy, time-locked to cue onset.

As can be observed in Figure 1, statistically significant decoding of the ERP waveform was achieved over the range 0 – 700 msec after cue onset. Notably, a small length of time before 0 msec also exhibits statistically significant above-chance decoding. It is most likely that this puzzling result is attributable to the temporal smearing of the low pass filter, by which the beginning of the data starting at 0 msec bled into the immediately preceding timepoints (Luck, 2014).

Experiment 2

ERP decoding of the cue period

The ERP decoding procedure that was applied to the EEG data from Experiment 1, described above, was also applied to the data from Experiment 2. Because Experiment 2 was a control experiment to account for the amount of above-chance decoding accuracy that could be expected solely on the basis of the visual system's sensory response to the physically different cue shapes, ERP decoding of the data from Experiment 2 provides analogous control over the results of the ERP decoding analysis of the data from Experiment 1. By comparing the decoding accuracy results of Experiment 2 with those from Experiment 1, the extent of the decoding accuracy purely attributable to the sensory responses to the cue shapes can be determined. Just as in the primary analysis, the a priori assumption is that information about the cue shape will have an impact early in the cue period, but any late-period EEG patterns will be dominated by cognitive processes, including attention.

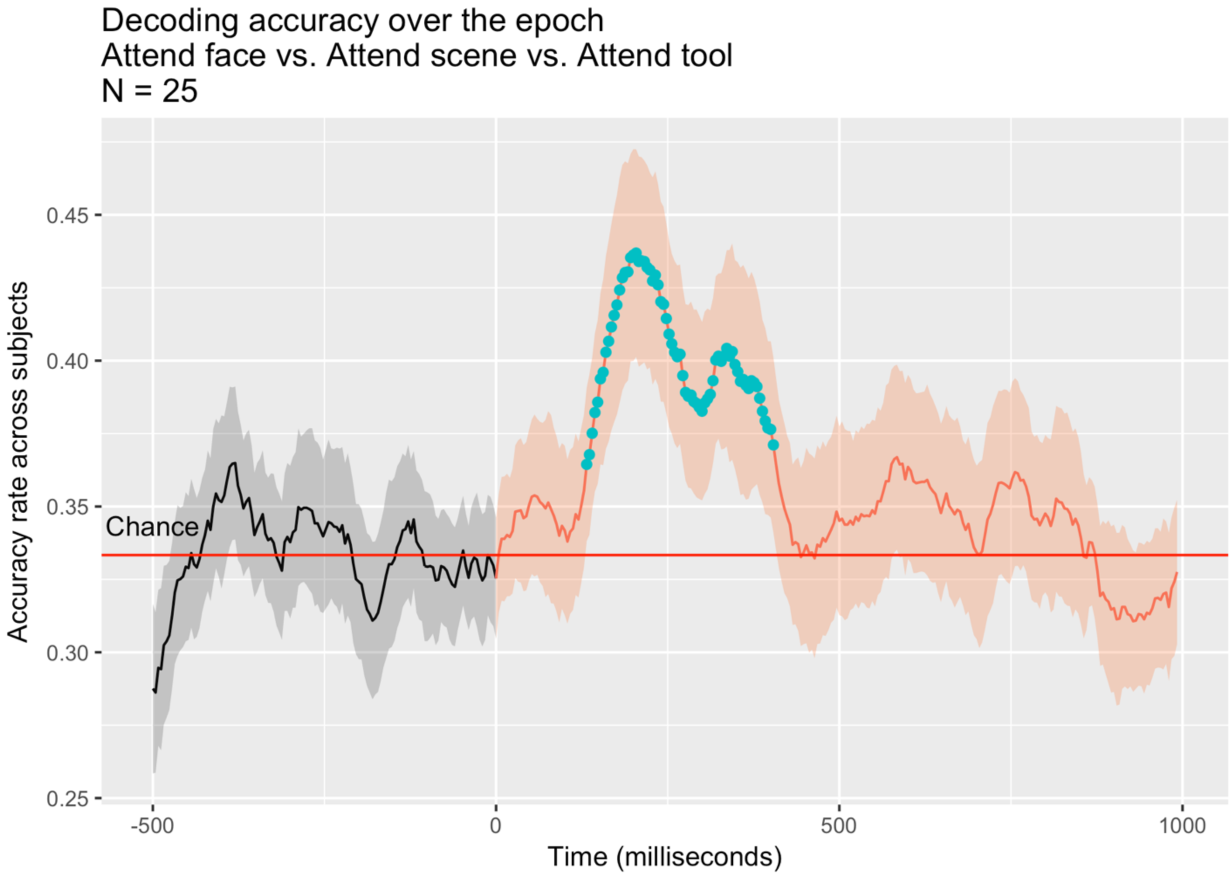


Figure 2. Experiment 2 ERP decoding accuracy, time-locked to cue onset.

From the decoding accuracy results displayed in Figure 2, it can be seen that the extent of the above-chance decoding accuracy is restricted to an early part of the cue period, rising sharply after the cue onset, rising again around the time that might be expected for the sensory response to the cue offset, and then ending before 500 msec after cue onset.

Experiment 3

ERP decoding of the cue period

The ERP decoding procedure that was applied to the EEG data from Experiment 1, described above, was also applied to the data from Experiment 3. Experiment 3 was a control experiment. In the primary analysis, the best case scenario for the decoding results of Experiment 3 were that they would replicate the decoding results of Experiment 1. In the case of the ERP decoding presented here, a similar logic applies. Theoretically, the SVM classifier should perform at chance level up until the cue onset time, because before the cue, there is no information available to the participants about the condition assignment for any given trial. After the cue, the same sensory response to the physical cue shape would be expected to drive decoding results in the early part of the cue period, and then later decoding accuracy would be attributable to cognitive processes and object-based attention. Because Experiment 3 controlled for the possibility that differences in task set across the three object conditions were driving the cognitive-period decoding results of Experiment 1, any above-chance decoding accuracy late in the cue period, after around 500 msec, would be attributable specifically to object-based attention.

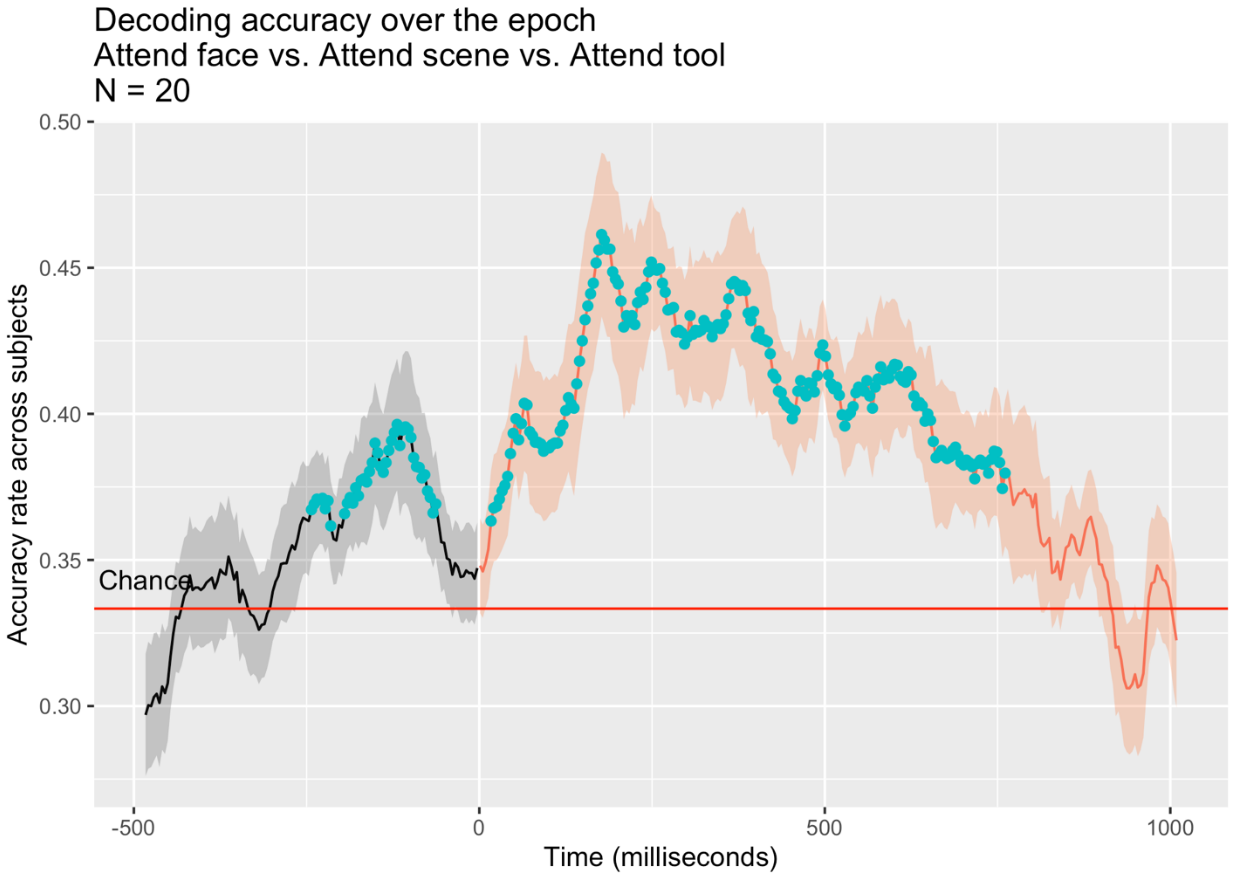


Figure 3. Experiment 3 ERP decoding accuracy, time-locked to cue onset.

As can be seen in the ERP decoding accuracy results for Experiment 3 presented in Figure 3, clusters of timepoints associated with above-chance decoding accuracy extend from shortly after cue onset into the late cue period, ending around 800 msec after cue onset. The results resemble those of the ERP decoding of Experiment 1 in that both experiments resulted in statistically significant periods of decoding later than would be expected for a purely sensory-driven effect.

ERP decoding of the target period

A nearly identical version of the ERP decoding procedure that was applied to the cue period EEG data from Experiments 1 – 3, described above, was also applied to the target period data from Experiment 3.

The participants' task in Experiment 3 was to press one button if the target object image, belonging to a cued object category, was blurry, and a second button if that target image was not blurry. On validly cued trials, target image stimuli were composite images, made of a randomly selected image from the cued object category overlaid with a randomly selected distractor image from one of the two uncued object categories. Both the target and the distractor object image could be blurry or not blurry, varying independently and randomly.

For the ERP decoding analysis presented here, the condition labels being decoded are blurry and not-blurry. The decoding routine is performed separately for the target and distractor objects in the stimuli, so that the effects of anticipatory object-based attention on stimulus representation can be compared for the attended object image and the unattended distractor object image. Invalid trials, in which the instructional cue did not predict the subsequently presented object image, were excluded from this analysis, because on invalid trials only one object image was presented in the stimulus: This object image did not belong to the cued object category, meaning that for invalid trial data, there is no available comparison between visual processing for attended and unattended object image information.

Because the information being decoded in this analysis is whether the stimulus image (target or distractor, depending on the analysis) is blurry or not-blurry, the only difference between the ERP decoding procedure applied here and the one detailed above for the ERP decoding analysis for Experiment 1 was that two condition labels were trained and tested, rather than three. Theoretical chance was thus set to 50%, both in the visualizations of results in Figure 4 and Figure 5, and in the cluster-based test for statistical significance.

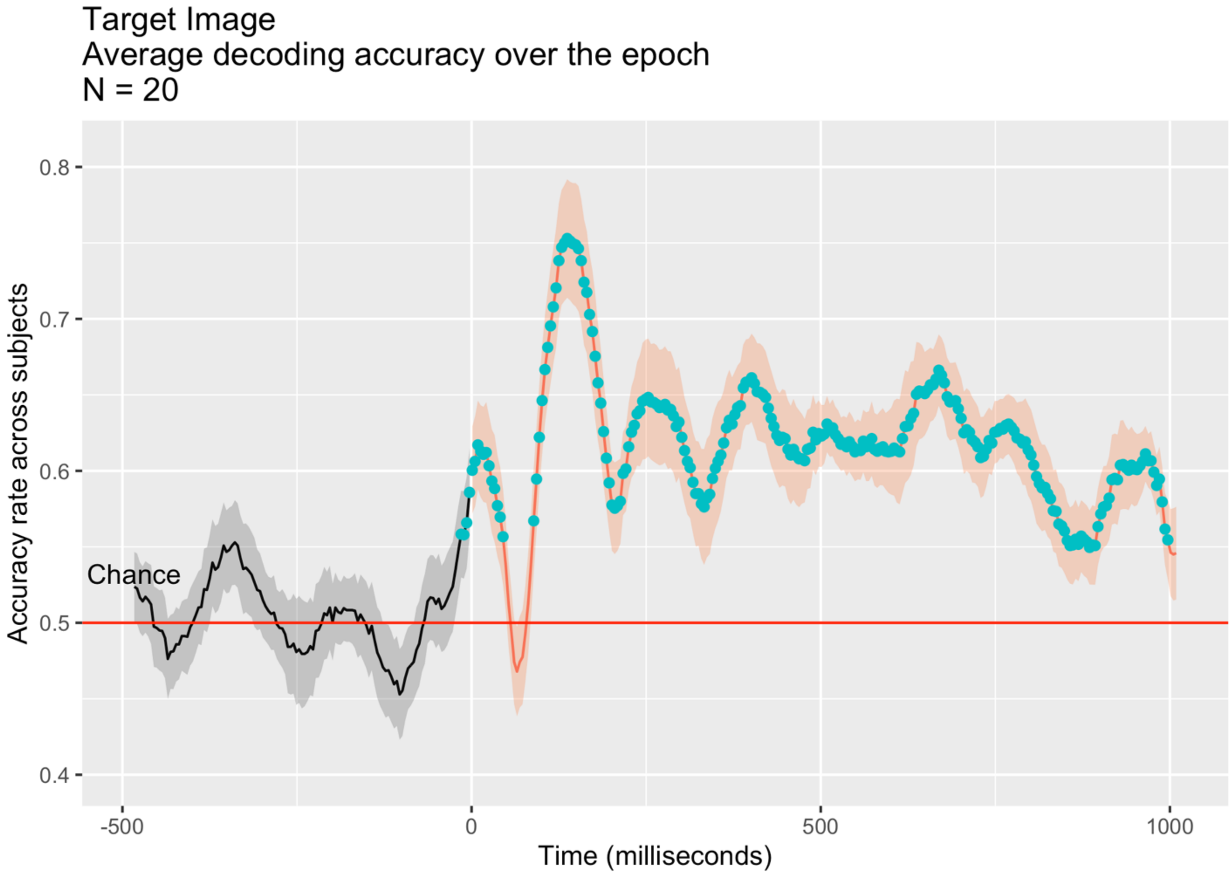


Figure 4. Experiment 3 target image blurriness decoding accuracy from the ERP, time-locked to target onset.

The ERP decoding results for the target object image are presented in Figure 4. Clusters of statistically significant time points persist from target stimulus onset to nearly the end of the decoding epoch. In the early sensory response period of roughly 0 – 200 msec, average decoding accuracy across participants peaks at 75%.

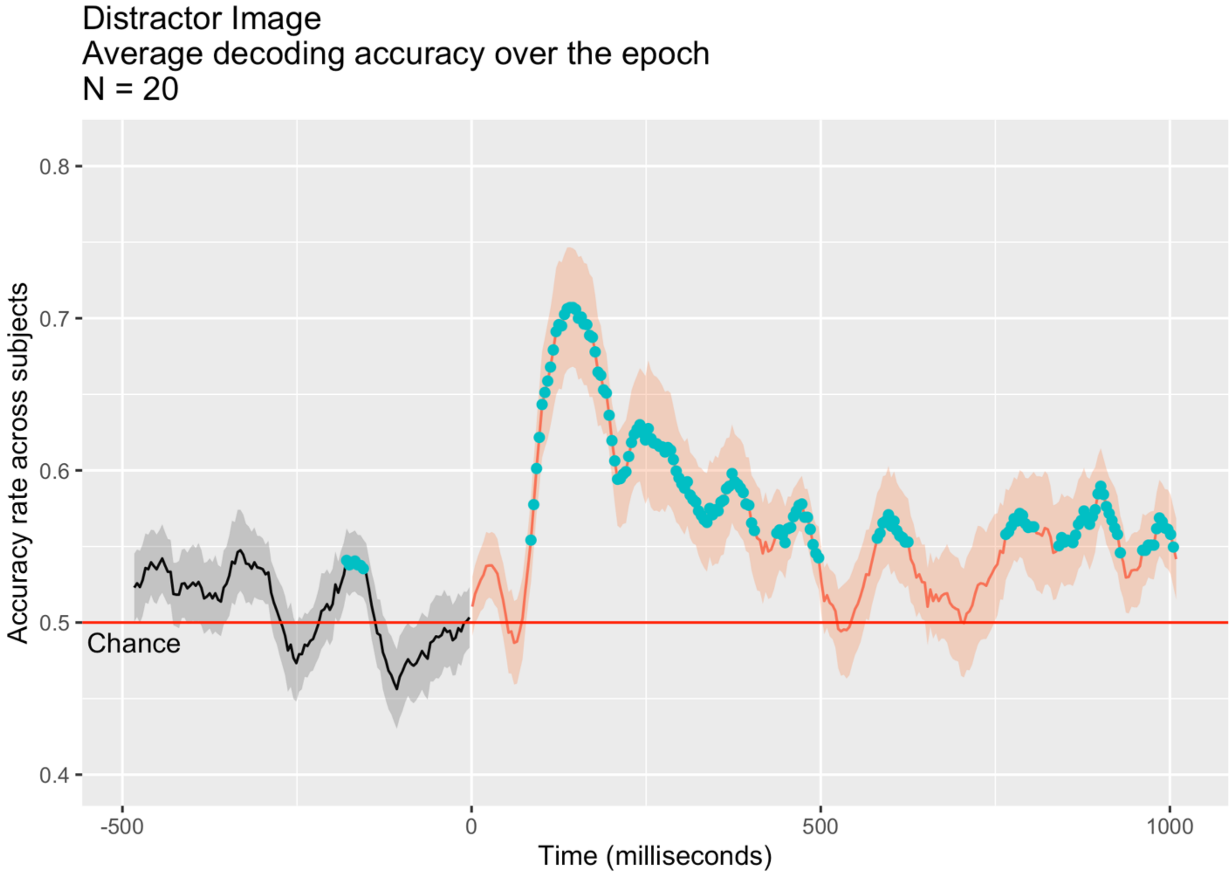


Figure 5. Experiment 3 distractor image blurriness decoding accuracy from the ERP, time-locked to target onset.

The ERP decoding results for the distractor object image are presented in Figure 5. Clusters of statistically significant time points begin at stimulus onset and decline after 250 msec, with small clusters of statistically significant decoding appearing later in the decoding epoch. In the early sensory response period of roughly 0 – 200 msec, average decoding accuracy across participants peaks at about 70%. Overall, compared to the target image decoding result, the decoding accuracy is weaker throughout the stimulus period.

The same ERP decoding analysis was also performed over Experiment 3 data segmented by whether the participants responded correctly or incorrectly. Responses were deemed correct when the blurriness or clarity of the cued object image in the target stimulus was indicated correctly by a button press during the interval between stimulus onset and cue onset of the subsequent trial. Responses were deemed incorrect when the response indicated by the button press during the response window did not match the blurriness of the cued object image – trials in which no response was recorded were excluded from this analysis.

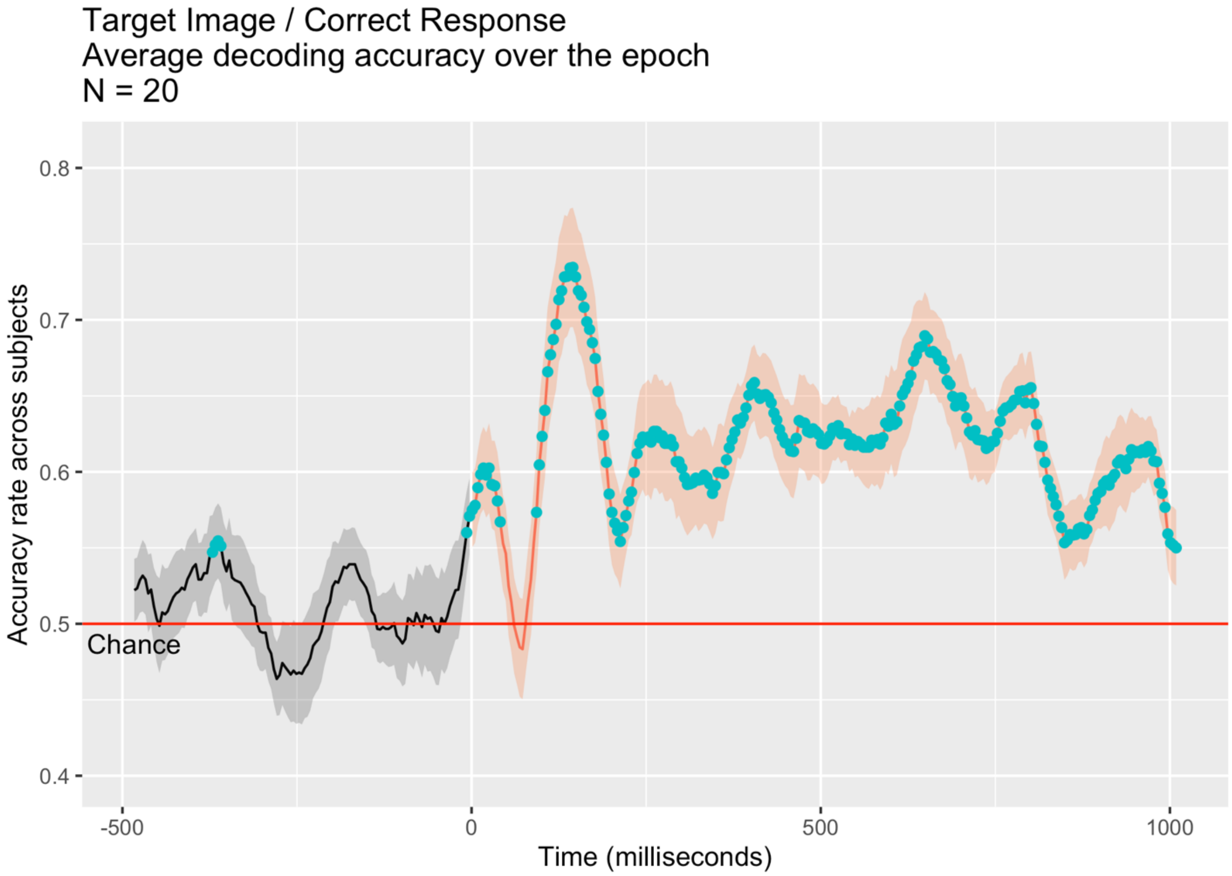


Figure 6. Experiment 3 correct-response trial target image blurriness decoding accuracy from the ERP, time-locked to target onset.

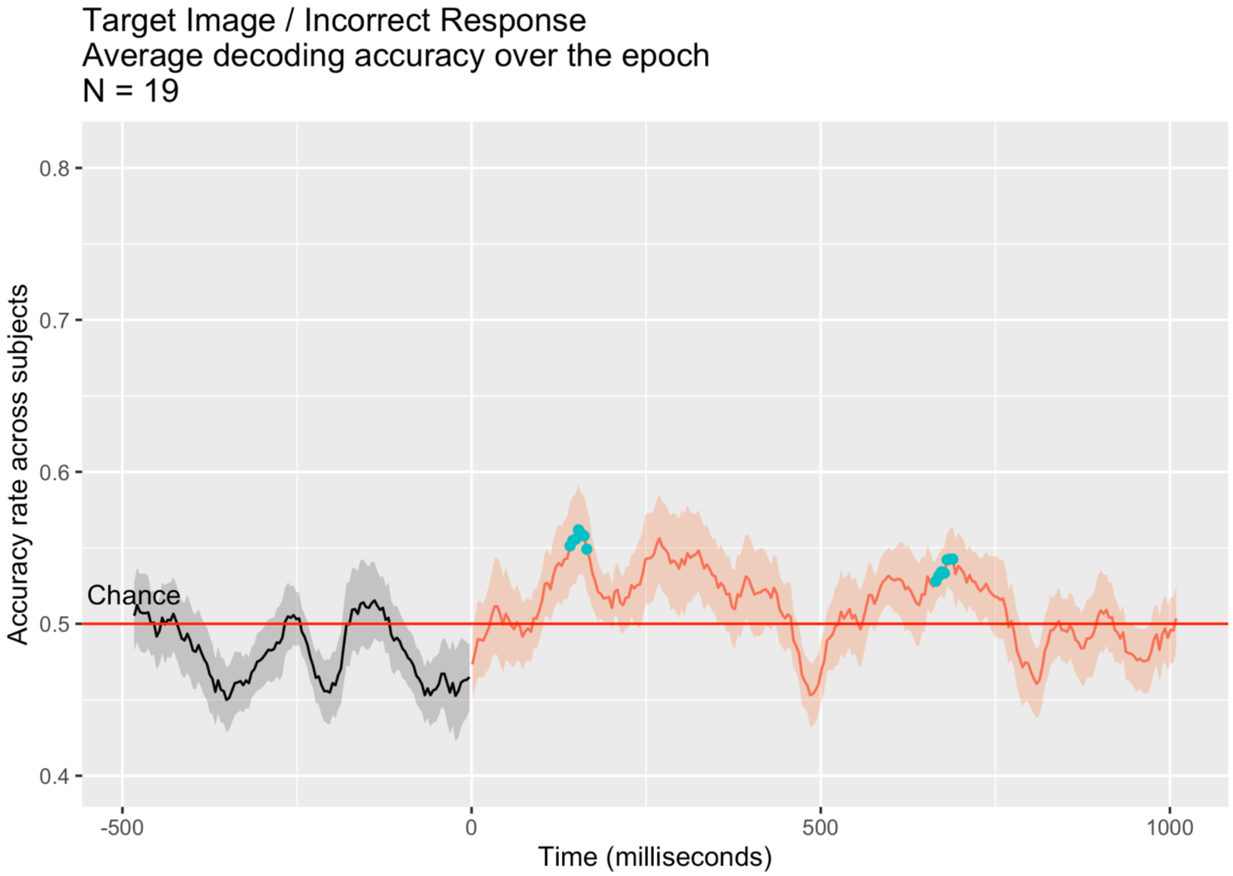


Figure 7. Experiment 3 incorrect-response trial target image blurriness decoding accuracy from the ERP, time-locked to target onset. The number of participants whose data is included in this analysis is 19, rather than 20, because one participant did not produce enough incorrect responses to meet the minimum requirements of the SVM cross validation procedure.

Figure 8. Experiment 3 correct-response trial distractor image blurriness decoding accuracy from the ERP, time-locked to target onset.

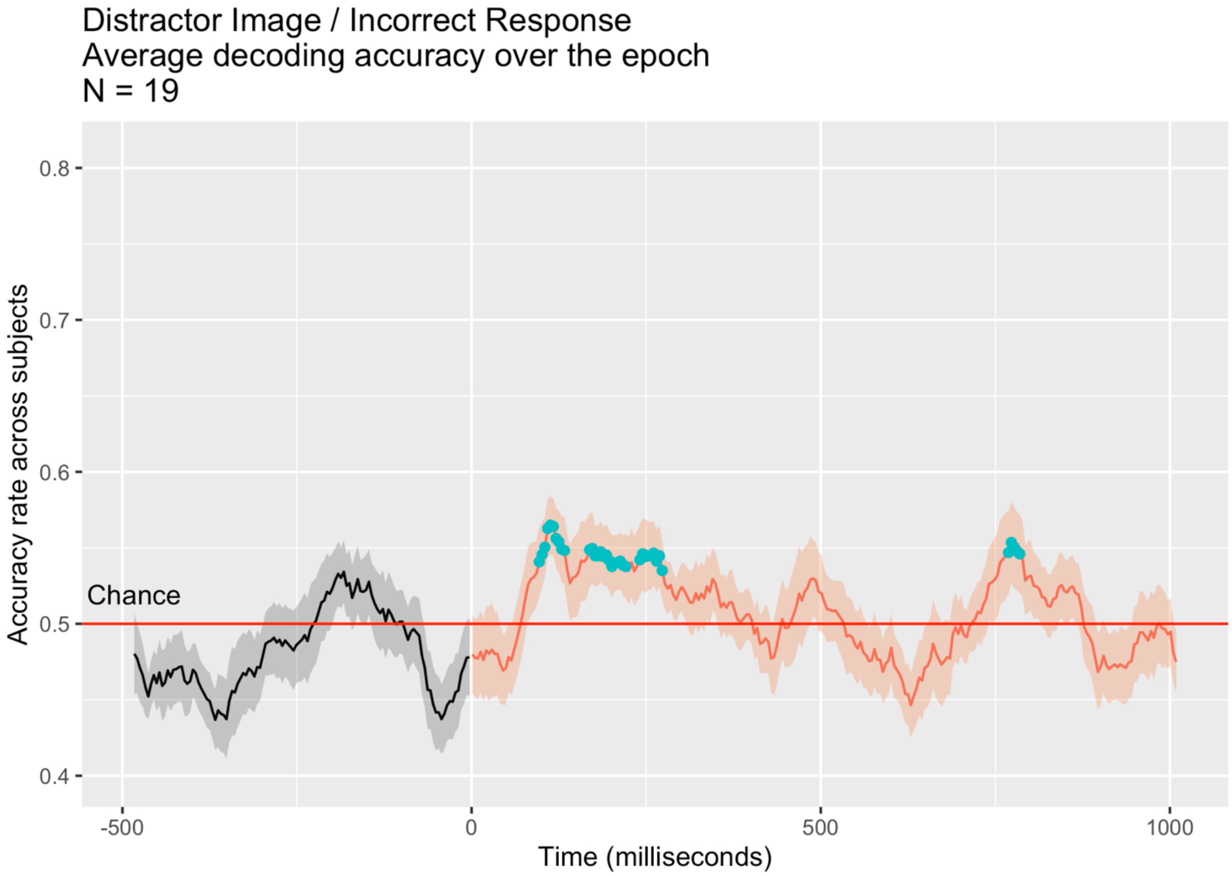


Figure 9. Experiment 3 incorrect-response trial distractor image blurriness decoding accuracy from the ERP, time-locked to target onset. The number of participants whose data is included in this analysis is 19, rather than 20, because one participant did not produce enough incorrect responses to meet the minimum requirements of the SVM cross validation procedure.

Comparison of cue period alpha decoding and target period ERP decoding

The primary analyses of Experiments 1 – 3 have provided evidence that patterns of alpha band power over the scalp are systematically modulated as a function of the category of object being selectively attended. An intriguing question follows from that series of findings: Does object category-specific alpha band modulation increase the amount of sensory information represented in the visual system from subsequently presented object images belonging to the cued category?

A key aspect of attention theory is that attention enhances the perception of target stimuli. According to early selection models, attention operates in the early stages of sensory processing, rather than only acting once all sensory information has been parsed and a semantic interpretation of the scene is available to higher levels of cognition and behavior. The early selection school of thought has largely been validated by experimental findings (Mangun and Hillyard, 1991), and thus any enhanced perception of targeted visual information afforded by attention should be accompanied by clearer and more distinct stimulus-specific sensory representations in cortex, relative to the cortical representations of unattended information. Furthermore, if alpha band modulation is a mechanistic element of selective attention, then the degree of alpha modulation should be directly linked to the sensory processing benefits afforded by attention. The design of Experiment 3 and the method of EEG decoding using an SVM classifier allows this prediction to be tested. Because the stimuli designed for Experiment 3 are composite overlays of both an image belonging to an attended object category and an image belonging to an ignored object category, and the blurriness of each object image in the overlay varies independently, the visual system's representation of visual information from each individual image can be directly compared. It is important to note, however, that this analysis is fundamentally exploratory in nature, because the hypothesis around which Experiments 1 – 3 were designed and executed did not pertain to this question of the direct relationship between alpha band oscillatory activity and visual perception. Nonetheless, the EEG data from Experiment 3 serendipitously allow for a test of this interesting question.

The analysis presented here allows us to examine whether anticipatory attention to an object category, mediated by alpha band modulation, increases the amount of visual information represented in neural activity elicited by a target object stimulus. Decoding accuracy can be interpreted as the minimum amount of information about the target contained in the decoded signal, so showing that object-based attention increases decoding accuracy of the blurriness of the attended object image but not of the blurriness of a simultaneously presented image from an unattended category would suggest that object-based attention acts at the level of a category representation to increase low-level sensory information pertaining to that category. A positive correlation of category-specific anticipatory alpha modulation with target representation enhancement would suggest that alpha activity plays a causal role in the perceptual benefit afforded by attention. Furthermore, a null or negative correlation between anticipatory period alpha band decoding and distractor decoding would also be expected. Whether the correlation between the alpha decoding measure and the distractor decoding measure is flat or negative implies one of two different theoretical interpretations, which will be elaborated in the following General Discussion chapter.

We ran two correlation tests: a between-subjects correlation of cue-period alpha decoding accuracy and target-period target visual information, and a between-subjects correlation of cue-period alpha decoding accuracy and target-period distractor visual information. The measure of cue-period alpha decoding accuracy we used was the peak decoding accuracy across the whole cue period, for each subject. The measure of target-period ERP decoding we used, for target object and distractor object visual information, was the peak decoding accuracy over the whole target-period, for the blurriness decoding of the respective image type, for each subject.

Figure 10. Between-subjects correlation of a measure of Experiment 3 cue period alpha band decoding accuracy and a measure of target period target image decoding accuracy from the stimulus-evoked ERP.

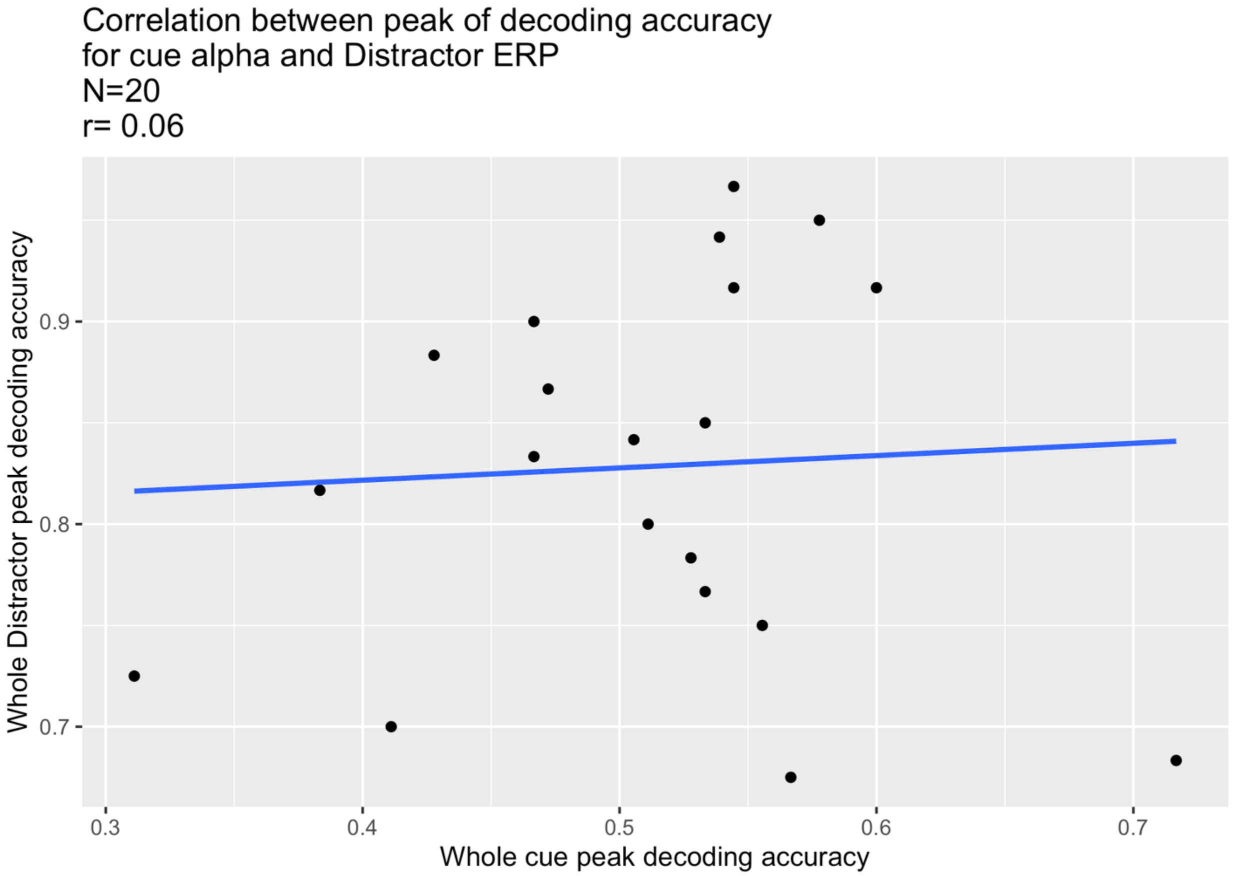


Figure 11. Between-subjects correlation of a measure of Experiment 3 cue period alpha band decoding accuracy and a measure of target period distractor image decoding accuracy from the stimulus-evoked ERP.

As can be seen in Figure 10, there is a positive correlation between our measure of EEG alpha power decoding accuracy for the cue period and ERP decoding accuracy for target object visual information in the target period. Figure 11 shows no correlation between the measure of EEG alpha power decoding accuracy from the cue period and ERP decoding accuracy for distractor object visual information in the target period.

Comparison of target and distractor decoding accuracy

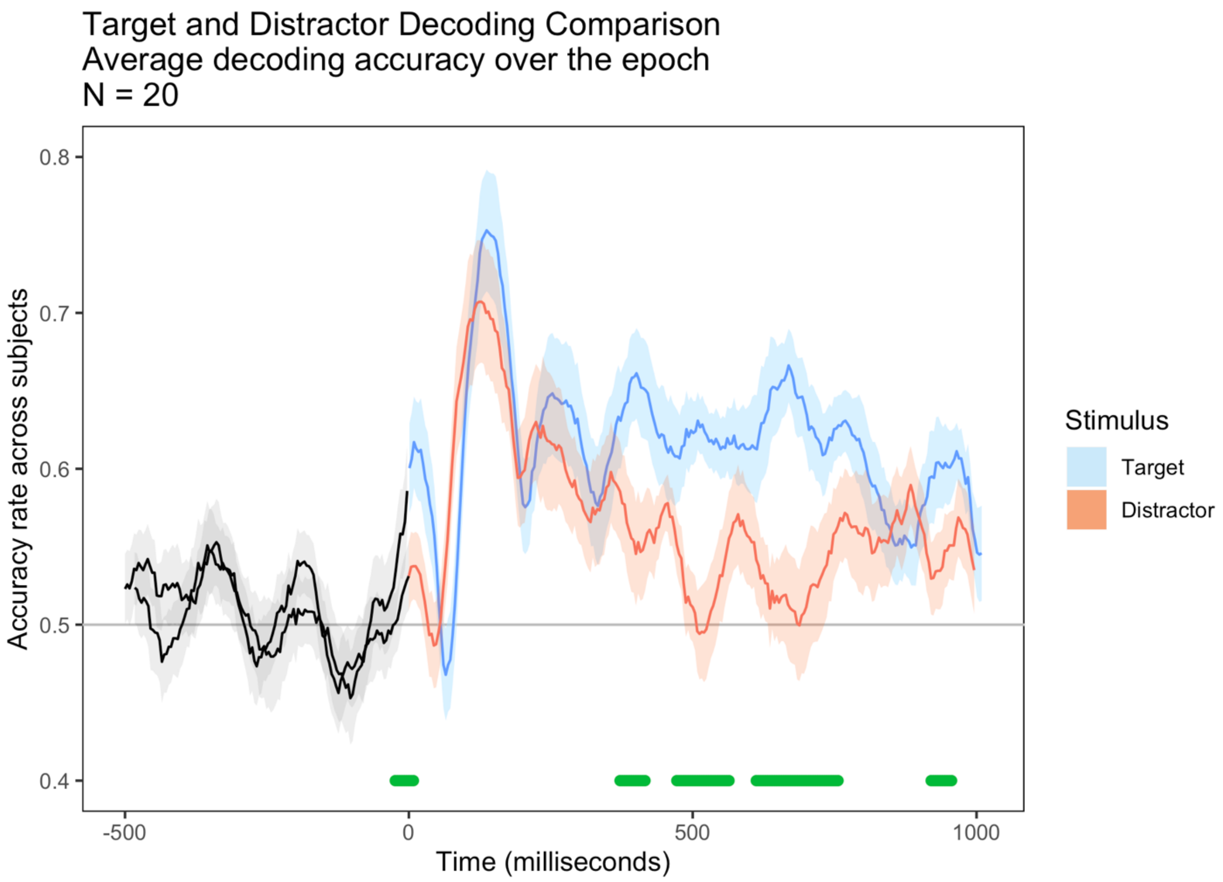


Figure 12. Comparison of Target and Distractor decoding accuracy. Target image decoding accuracy during the stimulus period, across subjects, is represented by the blue line and shaded region. Distractor image decoding accuracy during the stimulus period, across subjects, is represented by the orange line and shaded region. The grey horizontal line represents the baseline of chance-level decoding accuracy. The green lines at the bottom of the plot indicate time points belonging to clusters of time during which the decoding accuracies between Target and Distractor are significantly different, as indicated by cluster-based permutation test.

Using the same cluster *t*-mass permutation technique as described previously and utilized for all statistical significance tests of decoding accuracy against chance, statistically significantly greater decoding accuracy values were observed for Target decoding than for Distractor decoding. At each time point, the Target and Distractor decoding accuracy timeseries were compared by right-tailed paired-sample *t*-test to test the null hypothesis that target decoding was not more accurate than distractor decoding to a greater degree than might be expected by chance. The results of this statistical test are displayed in Figure 12. This test revealed statistically significant clusters of time points close to 0 msec (stimulus onset), and occurring several times throughout the period before 1000 msec.

Chapter 5: General Discussion

> *"The world is not so much made of stones as of fleeting sounds, or of waves moving through the sea." – Carlo Rovelli, The Order of Time*

Object-based attention is a fundamental component of natural vision. This assertion can be supported by two complementary lines of reasoning. First: People navigate the world principally on the basis of interactions with objects, which abound in typical environments (O'Craven et al., 1999; Scholl, 2001). The primacy of objects means that adaptive interaction with the world requires high-level object representations that are separate and distinct from all constituent low-level visual features in the same region of space. Therefore, an effect of attention directly on object representations is a critical aspect of perception (Woodman et al., 2009). Attention has been shown to operate on object representations (Tipper and Behrmann, 1996; Behrmann et al., 1998), so identifying the neural mechanisms by which attention influences object representations is a key goal in cognitive neuroscience. Second: The fact that the brain parses visual input into a panoply of objects – when in reality objecthood is a fabrication imposed on the external world rather than emerging from it veridically, and the world is actually a continually evolving and indivisible roil of matter and energy (Robinson, 1987; Rovelli, 2018) – indicates that objects are valuable cognitive constructs for the execution of adaptive behavior and survival of our species. The fundamental importance of object-based attention can be derived from this premise. This second line of reasoning cannot be easily operationalized in a cognitive neuroscience laboratory setting, so the first line of reasoning will be taken as the primary impetus for the experiments presently discussed.

Object-based attention is taxonomically subordinate to attention more broadly construed (Chun et al., 2011), and so an investigation into object-based attention's neural mechanisms should be situated within the larger construct of attention. Physiological studies show that attention's performance benefits correlate with neural activity changes in perceptual systems. Cortical structures coding attended information show increased signal amplitude, synchrony and/or functional connectivity (Moore and Zirnsak, 2017). How the nervous system mechanistically controls this cortical excitability and processing efficiency remains incompletely understood, but most models suggest that top-down control signals from higher-order networks in frontal and parietal cortex, such as the DAN, alter processing in sensory/perceptual cortical regions coding attended and unattended information (Petersen and Posner, 2012). One hypothesized neural signature of top-down control at the level of sensory/perceptual cortex is focal alpha power (Jensen and Mazaheri, 2010). Changes in alpha power occur during spatial attention (Worden et al., 2000), and feature attention (Snyder and Foxe, 2010a). These observed changes are proposed to reflect a mechanism of dynamically controlling information flow through the brain that has been called the Gating By Inhibition model (Jensen and Mazaheri, 2010). According to this model, alpha band activity is directly modulated by attention control regions in order to dynamically route the flow of feedforward information through sensory systems. This model has been supported by a study using a cued covert spatial attention design (Liu et al., 2016), which used simultaneous EEG-fMRI recording to link activity in the brain regions associated with the DAN. The Gating By Inhibition model is complementary to another model that ascribes functional importance to oscillatory neural activity: the Communication Through Coherence model (Fries, 2005). According to the Communication Through Coherence model, synchronization of spike firing in the gamma frequency band allows local neural populations to form an effective communication channel. Gating By Inhibition and Communication Through Coherence thus complement one another, because signal propagation by gamma band coherence can occur between the pulsed inhibitory phases of alpha band activity. Modulatory increases in alpha band power in a neural population thus shorten the windows of effective communication within and between that neural population and its upstream sources and downstream targets, and conversely, decreases in alpha band power enhance the population's information processing capacity (Jensen and Mazaheri, 2010).

Here we investigated alpha-based mechanisms mediating selective attention to objects by cuing attention to different objects and measuring changes in scalp-recorded EEG alpha power. This cued anticipatory object-based attention paradigm was designed to be analogous to the studies used to uncover alpha-based mechanisms in spatial and feature-based attention domains (Worden et al., 2000; Snyder and Foxe, 2010b). Altogether, we examined behavioral measures, EEG topographic maps, and machine learning classification results to test the hypothesis that object attention involves selective alpha power modulations in object-specific cortex.

We controlled for several possible alternative interpretations of our results, and performed ERP decoding analyses to corroborate our main findings and support our theoretical interpretation. Finally, in an exploratory analysis, we demonstrated that object-based attention enhances the visual system's representation of visual input specific to the targeted object category, and we identified a potential link between alpha band modulation and the amount of perceptual information represented in the visual system, further supporting the assertion that alpha band modulation is directly involved in the selective processing of sensory information.

The work discussed here consisted of three main experiments and a pilot experiment. Experiment 1 is the main experiment in which we tested whether EEG alpha band topographies could be differentiated between object-based attention conditions. Analysis of EEG data included topographic power difference map construction and SVM decoding of alpha band power to quantitatively assess whether the EEG alpha band contained information about the object category being attended.

The Pilot Experiment we initiated, but did not fully realize, was designed to control for the possibility that object-based attention was not necessitated by the design of Experiment 1. In the paradigm we designed for the Pilot Experiment, we sought to make engagement of object-based attention necessary for the perception of the target stimulus, which would be verifiable by chance-level performance on invalidly cued trials, in which the participant was instructed to engage anticipatory attention to a category of object that did not subsequently appear in the stimulus, and above-chance level performance on validly cued trials. However, the behavioral results we examined after collecting a preliminary dataset under this pilot design indicated that we did not actually achieve the cognitive manipulation that we set out to isolate, and so we continued our control experiments with two separate designs. Both of these subsequent designs resulted in full datasets.

In Experiments 2 and 3, we tested two alternative interpretations of our results from Experiment 1; one alternative interpretation was tested by each experiment.

Experiment 2 was designed to test whether decoding accuracy in the preparatory period between the cue onset and the target onset found in Experiment 1 might have been driven entirely by differences in the sensory processes evoked in the visual system by the different cue stimuli. The physical stimulus properties of the cues for the three different object attention conditions differed from one another (triangle vs. square vs. circle), and therefore, generate three different sensory responses in the visual system. It was likely that at least part of the decoding results we obtained from Experiment one was attributable to this confound, and it was conceivable, although theoretically unlikely, that our entire decoding result was attributable to this confound, rather than to the higher-level cognitive process of object-based attention, as we intended it to be. It was theoretically unlikely that this would be the case, because any systematic differences in sensory response should be restricted to the early period after cue onset and cue offset (Bae and Luck, 2018), and our statistically significant decoding result and decoding time period of interest was situated late in the trial epoch. Nonetheless, we conducted Experiment 2 to be able to confidently rule out the possibility that the cue shapes were driving our result, in lieu of being able to adequately counterbalance the cue shapes across participants and circumvent the potential confound in the design of Experiment 1.

In Experiment 3, we investigated whether differences in alpha topography across object attention conditions in Experiment 1 may have been the result of different task sets across the three object attention conditions, rather than reflecting object-based attention mechanisms in visual cortex. In Experiment 1, participants were instructed to press one button if the target image they saw was a male face, a natural scene, or a powered tool, and press a different button if the target image they saw was a female face, an urban scene, or a non-powered tool. The cue shape that preceded each target image instructed participants what object category to selectively attend in anticipation of the target display. Thus, depending on which object category was cued, it can be argued that the task set engaged by the participants was different for each type of trial – attend face, attend scene, and attend tool trials would effectively become discriminate male/female faces, discriminate urban/rural scenes, and discriminate powered/nonpowered tool trials. Conceivably, any decoding results we obtained in the anticipatory epoch might then reflect the differing task sets rather than object-based attention to the different object categories, per se. Experiment 3 was designed to circumvent this possibility that different task sets were driving the decoding result by equating the task across all three object categories. Instead of performing a subcategory discrimination as was instructed in Experiment 1, Experiment 3 instructed participants to simply discriminate whether the image of the cued object category was blurry or not blurry. Thus, in theory, the cue shape instructed participants to selectively attend to one of the three object categories, so that upon target presentation, the image belonging to the cued object category could be quickly and accurately parsed and the blurriness or clarity could be detected. So that participants were required to use object-based attention to perform this task, and not simply perform the task on the basis of a blurry-not blurry distinction, and thus attend to blurriness in the anticipatory period instead of attending to the instructed object category, the Experiment 3 stimuli were designed as composites of overlaid target and distractor object images, each of which could be blurry or not blurry, independently of the other.

We chose faces, scenes, and tools as attentional targets in Experiments 1 – 3 because these types of objects have been shown to activate circumscribed areas in the visual cortex. The fusiform face area (FFA) is selectively responsive to images of upright faces (Allison et al., 1994; Kanwisher and Yovel, 2006): Faces can be considered objects because, for example, evidence from patients with prosopagnosia suggests that the similar mechanisms underlie face recognition and object recognition (Gauthler et al., 1999). The parahippocampal place area (PPA) is responsive to scenes (Epstein et al., 1999), and specifically to scene category (Henriksson et al., 2019). Areas responsive to tools have been identified in the ventral and dorsal visual pathways (Kersey et al., 2016). In line with the prediction that object-based attention modulates alpha in visual areas specialized for processing the attended object category, attention to faces should selectively decrease alpha band activity in face-selective visual areas like FFA, attention to scenes should decrease alpha band activity in place-selective areas like PPA, and attention to tools should decrease alpha band activity over tool-selective regions of the ventral and dorsal visual pathways. EEG is not a strong method for localizing the neural sources of brain activity, but given that the FFA, PPA, and postulated tool-specialized areas are located in different cortical regions, the patterns of alpha modulations with attention in these areas would be expected to produce differential EEG alpha patterns on the scalp. Given that such patterns might be expected to be only subtly different, and in ways difficult to predict, one avenue for assessing different patterns of alpha for attention to different objects is to incorporate machine learning to decode scalp EEG alpha patterns. Such differences should only be expected if focal modulation of alpha is also involved in selective object attention.

Our reaction time results showed that participants engaged object-based attention to cued object categories, being faster to identify cued objects. Theoretically, when cued to anticipate a particular object category, participants would bias neural activity within the cortical area(s) specialized for that object type, and perhaps also bias activity within cortical areas processing all the lower-level visual features associated with that object (Cohen and Tong, 2015). When the target appears, its visual properties would thus be integrated, facilitating the required perceptual discrimination. When the object appearing is from an unanticipated (uncued) category, activity in object selective areas and associated visual feature areas for the uncued objects would be relatively suppressed, delaying the integration and semantic parsing of uncued target images, and slowing reaction times. This pattern of reaction time difference between valid and invalidly cued trials was observed in both Experiment 1 and Experiment 3. Because Experiment 2 was not designed to engage anticipatory object-based attention, there was no operationalization of attention, and no difference between attention conditions to examine with behavioral measures.

Topographic alpha power difference maps varied with the object category that was attended. Differing alpha power topographies were consistent with scalp EEG patterns that would be expected if the alpha modulations were occurring in different underlying cortical generators (cortical patches or areas) for the three object categories. The wealth of evidence about underlying neuroanatomical substrates of face, scene, and tool processing from imaging studies allows some predictions about our data with respect to the hypothesized nature of the focal cortical activity contributing to our topographic and decoding findings. The right hemisphere-emphasized FFA (Kanwisher et al., 1997), and the equally bilaterally distributed PPA (Epstein and Kanwisher, 1998), would, in principle, predict a differential scalp alpha distribution, and perhaps lower alpha power broadly over the right occipital when attending faces. Our attend-face minus attend-scene alpha topography was generally consistent with this prediction (Chapter 4; Figure 3A), and this pattern was different from that in the attend-face minus attend-tool difference plot (Chapter 4; Figure 3B). We hope to make exceptionally clear, however, that we are <u>not</u> proposing that we can localize the underlying cortical generators of scalp-recorded activity using the methods we employed here. Furthermore, the representation of objects over distributed networks of neural populations in IT cortex, detailed in the Introduction chapter of this dissertation, suggests that although there are relatively circumscribed visual areas that are responsive to images from specific object categories, there is also significant potential cortical overlap between these object-specific networks, and any subtle differences in object category-respective network activity are far beyond the limit of what EEG can distinguish. For these reasons, we turned to SVM decoding for our principal measure of object-based attention's effect on alpha power modulation.

Our decoding analyses provide strong support for the claim that attention modulates alpha topographies in an object category-specific manner, and is in line with the time courses of the differences in alpha patterns observed in the scalp topographic difference plots. In our decoding analyses, statistically significant above-chance decoding accuracy provides straightforward evidence that alpha topography contains information about the selected object category, and therefore, that top-down object-based attention modulates alpha topography according to the cued (attended) object category. We observed that statistically significant decoding occurred in the 500 – 800 msec range post-cue/pre-target, indicating that patterns of alpha topography at the scalp were reliably modulated by our attention manipulation in this time range (Chapter 4; Figure 4). Importantly, the 500 – 800 msec range corresponds to the periods in the alpha topographic difference plots where the patterns stabilized.

In order to test whether our decoding results were specific for the alpha band, we performed the same SVM decoding routine on theta, beta, and gamma band power and found no significant above-chance decoding in the anticipatory period for those frequency bands (Chapter 4; Figure 5). This result is consistent with the hypothesis that oscillatory neural activity in the alpha band is mechanistically involved in anticipatory attention, whereas activity in other EEG frequency bands is not modulated in target-relevant visual areas in human EEG.

In our two follow-up experiments, we directly assessed the two alternative interpretations of our decoding results from Experiment 1, described in more detail above. In brief, the first confound examined was whether differences in alpha scalp topography post-cue might reflect purely sensory processing associated with each cue (e.g., triangle vs. circle). This should be applicable only to the above-chance (although not significant by our tests) decoding observable in the early post-cue period (~ 0 – 200 msec) of our decoding results, and not the significant longer-latency decoding. Indeed, we verified this in Experiment 2, in which participants performed the same task, and saw the same cues and targets as in Experiment 1, but the cue shape did not predict the upcoming object category. We observed statistically significant decoding in the post-cue/pre-target period from 0 – 200 msec attributable to physical cue features (e.g., Bae and Luck, 2018), but no significant decoding later in the cue-to-target interval.

The second alternative explanation of our decoding results from Experiment 1 is that they were driven by task set differences across cued object conditions. The task for faces, for example, was to discriminate gender, while for scenes it was to distinguish between urban scenes and natural scenes, leaving open the possibility that our decoding late in the post-cue period reflected task set differences (Hubbard et al., 2019) rather than attentional control over object selection as we propose. We can reject this alternative based on the results of Experiment 3, in which the cues predicted the relevant target object, but the discrimination task was the same for all object categories

– discriminate whether the cued object was in focus or blurred. We were thus able to replicate the longer-latency alpha-related preparatory attention effects reported in Experiment 1 while controlling task set factors.

Our findings show that EEG alpha modulation is linked to object-based selective attention, extending previous findings that alpha modulation is associated with attention to spatial locations and low-level visual features. Using an SVM decoding approach we identified differences in the topographic patterns of alpha power during selective attention to different object categories. Further, we linked the time range during which statistically significant decoding was achieved to alpha power topographic maps, and observed that alpha modulation was consistent with the time course of preparatory attention observed in prior research. Overall these findings support the model that alpha band neural activity functions as an attentional modulator of sensory processing for both low level visual features and high-order neural representations such as those for objects. Our results add to the body of theoretical and experimental work suggesting that alpha band modulation is an elemental part of the neural mechanisms of attention. We can now demonstrate empirically that for at least three distinct levels of visual processing – spatial organization, feature organization, and object representations – alpha band modulation is associated with selective attention. This finding is consistent with the SpOC model of attentional control, to the extent that the model postulates an identical mechanism of attentional enhancement at any sensory site that can be the target of attentional modulation.

ERP decoding analyses conducted over the EEG datasets from Experiments 1 – 3 support the basic assertions of our theoretical interpretation, and go even further in establishing a direct link between attentional alpha band modulation and the immediate visual information available to perception.

First, we performed SVM decoding over ERP waveforms time locked to the cue onset. The results of these ERP decoding analyses corroborate the alpha power decoding results. Experiment 1 and Experiment 3 exhibit similar decoding results, with large above-chance peaks early in the cue epoch, and sustained levels of above-chance decoding accuracy later into the cue period. That the results of these two experiments display similar patterns of ERP decoding strengthens our interpretation of the primary analyses of the experiments. We took the alpha band decoding results of Experiment 1 to mean that object-based attention, directed selectively to a single object category, altered the pattern of alpha band activity throughout the visual system as a function of the targeted object category. Because the alpha band decoding results of Experiment 3 resembled those of Experiment 1, we concluded that object-based attention, and not a different cognitive process, like task set or response mapping, or a set of different processes, was driving the systematic alpha power distributions. Because we replicated this basic decoding result with the ERP data, and because the ERP data that was fed into the SVM decoding analysis was filtered to the frequency band 0 – 6 Hz so that it would contain no overlapping information with the alpha frequency band used in the primary set of analyses, we have further evidence that the same cognitive process was engaged during the cue period in both Experiment 1 and

Experiment 3. If the ERP decoding results from Experiment 3 did not resemble those of Experiment 1, then we would be forced to call into question our interpretation that the similar patterns of alpha decoding results that we observed were caused by the same underlying cognitive processes. Instead, the replication provides corroborating evidence that the same engagement of object-based attention was induced across the two experiments.

We then performed the same kind of ERP decoding analysis over the target-period EEG data from Experiment 3. The stimulus design of Experiment 3 allowed for a test of the perceptual benefits afforded by object-based attention, using the decoding method to obtain an estimate of the image information represented in the EEG data, and thus in the underlying cortical activity elicited by the stimulus presentation and the cognitive context. Because the Experiment 3 stimuli were composite images made of an image from a target object category overlaid with an image from a distractor object category, the amount of perceptual information belonging separately to the target and the distractor object image in the same physical stimulus can be assessed. For the present decoding analysis, this perceptual information took the form of whether the object image was blurry or clear. We performed an SVM decoding analysis on the ERP waveforms elicited by stimulus images with two classification labels: blurry and not-blurry. We performed the same decoding analysis twice: once with respect to the target (cued) object images in the composite stimuli, and once with respect to the distractor (uncued) object images in the composite stimuli. Invalidly cued stimuli were excluded from this analysis, because those stimuli only contained one object image (overlaid with a distracting checkerboard pattern). The same physical stimuli were included in both the target decoding analysis and the distractor decoding analysis, and only the psychological condition of object-based anticipatory attention varied between the two analyses. Therefore, this analysis gains credence by adhering to the Hillyard Principle of experimental design (Luck, 2014).

The analysis comparing target and distractor decoding revealed significant differences in the decoding accuracy at multiple times in the stimulus period (Chapter 4; Figure 12). Because the two decoding analyses were performed over the same stimuli, it can be inferred from the decoding results that object-based attention increases the amount of visual information belonging to the attended object image that is represented in cortex, even very early after stimulus onset. This inference is intriguing because it suggests that attention to a specific category of object, engaged in anticipation of the presentation of that object, causes preferential processing and representation of low-level visual information consistent with that object category. The abstract form of the object is in some way instantiated prior to its eventual appearance, and this influences the low-level visual processing of features that fit into that abstract form. This interpretation illuminates the time course of the perceptual effects of object-based attention, and situates our object-based attention study within a larger literature on early selection models of attention.

On its face, the difference in decoding performance later in the target period, with the target decoding outperforming the distractor decoding throughout most of the stimulus period, cannot safely be attributed to perceptual differences arising from object-based attention, because the decoding analysis is confounded with the behavioral response instructions. The task in Experiment 3 was to press one button if the cued image was blurry and a different button if the cued image was not blurry. Therefore, it cannot be determined whether the decoding results in the target period past the initial perceptual phase are attributable to the maintenance of target image visual information, or response preparation and execution (index finger movement versus middle finger movement), or some combination of these two potential drivers.

In order to strengthen the interpretation that the differences in decoding accuracy between attended and unattended object images over the stimulus period were driven by object-based attention and not motor response preparation, we repeated the decoding analysis after segmenting the data according to whether it belonged to a correct-response or an incorrect-response trial. We observed that for incorrect-response trials, decoding the blurriness of the target object image performed much worse than decoding the target object image blurriness for correct-response trials: The long tail of statistically significant above-chance decoding observable in the decoding accuracy for the correct-response trials is not present for the incorrect-response trials. This finding suggests that the persistent representation of information about the attended object image throughout the decoding epoch is driven by attention, and not response preparation. On incorrect-response trials, motor processes are still mapped to blurriness conditions (but the mapping is the reverse of the instructed mapping), and thus motor response is just as confounded with target blurriness as for correct-response trials, but the fact that the participants responded incorrectly indicates that they did not engage object-based attention sufficiently strongly in the anticipatory period, and therefore were not able to perceive the targeted object image with enough fidelity to detect whether it was blurry or clear. The near-chance level performance of incorrect-response trial decoding thus places an upper limit on the decodable differences in EEG signals from index finger versus middle finger response preparation and execution, and makes attribution of correct-response decoding results to confounded motor response less plausible. This follow-up analysis, breaking down the Experiment 3 data by response accuracy, thus strengthens the case that object-based attention enhances the representation of visual information during stimulus duration and all the way through to behavioral response planning and execution.

Lastly, we examined the relationship between cue-period alpha decoding and target-period ERP decoding. Theoretically, if the SpOC model is accurate in its depiction of a common attentional enhancement mechanism imposed on sensory sites via signals issuing from a top-down control hub, then the degree of differential alpha modulation induced by anticipatory object-based attention should correlate with the degree of attention's benefit to perception. The degree of differential alpha modulation can be indexed with our SVM alpha power decoding method. If anticipatory attention targeting face images alters the visual system-wide pattern of alpha band activity such that areas selectively responsive to face information have less alpha power than other areas, and alters the pattern of alpha analogously for scene and tool attention, then there should be distinguishable patterns of alpha band power on the scalp. This conjecture is the logic underlying our use of alpha power topography SVM decoding. A corollary inference that follows from this logic is that ineffective attentional inducement of systematic alpha band activity patterns, or lack of attention either by task noncompliance or individual attentional deficit, would result in similar or identical patterns of alpha power over the scalp across all object attention conditions. The more similar the alpha power topographies are by object condition, the less information the SVM learner has for classification, and therefore the worse its decoding performance. In this way, any individual participant's alpha SVM decoding performance indicates the extent to which their anticipatory attention is selective for a particular object category.

On the basis of this line of reasoning, we predicted that we would see a positive between-subjects correlation between cue-period alpha decoding and target-period ERP decoding for target object image blurriness, and we would see a negative or null correlation between cue-period alpha decoding and target-period distractor image ERP decoding. In other words, we predicted that the more precisely and effectively an individual participant could engage object-based attention in an anticipatory manner, the better their perception of the stimulus image belonging to that object category. Their perception of the distractor image category could be either worsened by focused attention away from that object category, which would be observable as a negative correlation between cue-period alpha decoding and target-period distractor decoding, or unaffected, which would be observable as a flat correlation between cue-period alpha decoding and target-period distractor decoding.

In our between-subjects correlation analysis we observed the predicted outcome. We found a positive between-subjects correlation of the peak alpha decoding accuracy in the cue period and the peak ERP decoding accuracy for target blurriness in the target period. The correlation did not reach statistical significance ($p = 0.08$), but this finding provides new evidence to suggest that alpha band modulation is a component of the neural mechanisms of object-based attention, and visual attention more broadly. Moreover, we found a null between-subjects correlation of cue period alpha decoding and distractor ERP decoding. This finding suggests that the effective anticipatory selection of a cued object categories may leave unaffected the representation of visual information from an uncued, distracting object category. Together, these findings provide evidence for a causal functional role of alpha band oscillatory activity in the implementation of selective object-based attention. Taken to their logical conclusion, it may then be hypothesized that top-down object-based attentional modulation might act via the local suppression of alpha band activity in cued object-selective neural populations, fitting with the observed positive correlation between cue-period alpha decoding and stimulus-period target decoding and null correlation between cue-period alpha decoding and stimulus-period distractor decoding. However, establishing this mechanism rigorously will require further study.

Chapter 6: Conclusion

What are the most important points to take away from this dissertation? This concluding chapter will recapitulate the theoretical groundwork laid out in the Introduction chapter, and then will also summarize the original empirical results that are significant because of how they fit into the larger context of the cognitive neuroscience of attention. Finally, future directions indicated by this research will be described.

The primary theoretical groundwork that supported the original research described in this dissertation is the conceptual development of a potential system for routing the flow of information through sensory cortex based on the dynamic control of oscillatory neural activity. Gating by oscillatory dynamics would represent a functional mechanism for dynamically routing information flow. This functional mechanism possesses an advantage over anatomical mechanisms, the speed of which is limited by the kinetics of vesicle release and synaptic plasticity, and thus not conducive for operation at the speed of cognition and behavior. In line with this theory, attentional selection mechanisms in visual cortex have been found to involve changes in oscillatory activity in the EEG alpha band (8 to 12 Hz) – with decreased alpha indicating focal cortical enhancement and increased alpha indicating suppression. This pattern has been observed for spatial selective attention and attention to stimulus features such as color versus motion.

In three experiments, we investigated whether attention to objects involves similar alpha-mediated changes in focal cortical excitability. Our experiments utilized a cued object-based attention design, in which participants were instructed to engage selective attention to one of three specific object categories at a time, in order to enhance their perception of subsequently presented object images. Support vector machine decoding of alpha power patterns obtained during the anticipatory period revealed that late (>500 msec latency) in the cue-to-target foreperiod, EEG alpha indeed differed as a function of the to-be-attended object category. In Experiment 2, to eliminate the possibility that decoding of physical features of the cues led to our results, we designed a similar task, with the crucial difference that cues were non-predictive of the object category. There was thus no engagement of object-based attention in this control experiment. Alpha decoding was now only significant in the early (<200 msec) foreperiod. In Experiment 3, to eliminate the possibility that task set differences between the different object categories led to our Experiment 1 results, we designed a predictive cuing task where the discrimination task for different objects was identical across object categories. The results replicated Experiment 1. Together, these findings support the hypothesis that the neural mechanisms of visual selective attention involve focal cortical changes in alpha power for not only simple spatial and feature-based attention, but also high-level object attention in humans, indicating a common mechanism of attention throughout the visual system.

We followed up these primary analyses with ERP decoding analyses that strengthened our interpretation of the results from Experiments 1 – 3. Furthermore, by decoding target image properties from Experiment 3 EEG data, we observed that object-based attention directly increased the amount of information about the target object image that is represented in cortical activity, relative to the amount of information represented from simultaneously visible stimuli.

Ultimately, I hope that the question that motivated our experiments – whether oscillatory neural activity in the alpha band is modulated as a function of object-based attention – is sufficiently answered by the series of experiments described in this dissertation. I also hope that the larger significance that our experimental results point to – that alpha band modulation is a general mechanism of selective attention throughout the visual system – is made appreciable by the theoretical context expounded throughout this document. One final reason that this work is a significant piece of the encompassing scientific puzzle is that it seeds intriguing new questions for future experiments to answer, and thereby propagates the continuing scientific enterprise of understanding how attention is implemented in the brain. Looking forward, the original experiments described in this dissertation set the stage for future work to strengthen our theoretical interpretation and to further develop the SpOC model of attention control.

One of the most obvious questions that remains to be answered is whether the modulation of alpha band activity can be localized to the visual areas that they should theoretically inhabit. Theoretically, the power of alpha band activity should be relatively lower in the visual areas that are selectively responsive to objects targeted by attention than visual areas responsive to other categories of objects. In the case of the object stimuli utilized in Experiments 1 – 3 of this dissertation, these areas include the fusiform face area, the parahippocampal place area, and areas of the ventral and dorsal visual pathways, for face, scene and tool stimuli respectively. Future work utilizing simultaneous EEG-fMRI recording could be suited to answer the question of whether alpha band activity is focally modulated in these areas as would be predicted by theory, following the strategy of Liu and colleagues (Liu et al., 2016).

Another important subject to investigate is the involvement of a central, top-down control area in object-based attention. According to the simplest version of the SpOC model, the same network of frontal and parietal areas issues control signals to sensory sites regardless of the type of attention deployed – spatial, feature-based, or object-based. The SpOC model predicts that the fine structure of this network, or differentially activated subnetworks, contain the map of different sensory areas that can be selectively targeted. This prediction can be tested with an fMRI protocol that uses MVPA and decoding to measure how much information about attentional instructions can be represented within the DAN as a specified region of interest.

Following from this future line of work would be the further question of how exactly signals issued from the attention control source in the DAN mechanistically modulate alpha band activity in sensory sites. This question might be addressed with a computational modeling study that explores the tolerances of different oscillatory firing regimes in biologically plausible neural network models of sensory cortex, and identifies perturbations that when paired with given initial conditions in the system can initiate synchronized oscillatory firing patterns or break the network out of these patterns. Any such perturbations might then be sought out biologically in the top-down signaling from the DAN or its relay connections to sensory cortex.

These prospective lines of inquiry are all made especially interesting in light of the results of the experiments described in this dissertation. That oscillatory neural activity in the alpha band is elemental to the implementation of attention uniformly throughout the visual system points the way to specific, testable questions that can drastically improve our understanding of attention in the brain.

And understanding how biological organization and neural architecture support attention would represent a major step toward the overarching goal of explaining how the rote material processes of the physical universe are linked to the subjective interiority of the mind. Attention is a necessary component of conscious awareness (Noah and Mangun, 2020). Investigating the physical implementation of attention is thus a wedge into the empirically unbreachable problem of understanding the physical implementation of consciousness in itself. The hard problem of consciousness is both one of the most important and the most seemingly insurmountable problems faced by science (Chalmers, 2018), and so the possibility that consciousness can be effectively pinned down by its intrinsic link to the operationalizable phenomenon of attention hints at a tenable path toward scientific comprehension of this singularly vexing subject. Whether this path will lead to a satisfying theory of reality and our place in it is unknown, but the journey is enticing nonetheless.

> *"Science cannot solve the ultimate mystery of nature. And that is because, in the last analysis, we ourselves are part of nature and therefore part of the mystery that we are trying to solve." – Max Planck, Where is Science Going?*

References

Allison T, Mccarthy G, Nobre A, Puce A, Belger A (1994) Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. Cereb Cortex 4:544–554.

Bae GY, Luck SJ (2018) Dissociable Decoding of Spatial Attention and Working Memory from EEG Oscillations and Sustained Potentials. J Neurosci 38:409–422 Available at: https://www.ncbi.nlm.nih.gov/pubmed/29167407.

Baldauf D, Desimone R (2014) Neural Mechanisms of Object-Based Attention. Science (80-) 1268:424–428.

Behrmann M, Zemel RS, Mozer MC (1998) Object-Based Attention and Occlusion: Evidence from Normal Participants and a Computational Model. J Exp Psychol Hum Percept Perform 24:1011–1036.

Berger H (1929) “Über das Elektroenkephalogramm des Menschen.” Arch Psychiatr Nervenkr.

Bertschinger N, Natschläger T (2004) Real-time computation at the edge of chaos in recurrent neural networks. Neural Comput 16:1413–1436.

Betti V, Corbetta M, de Pasquale F, Wens V, Penna S Della (2018) Topology of functional connectivity and hub dynamics in the beta band as temporal prior for natural vision in the human brain. J Neurosci.

Biederman I (1987) Recognition-by-Components: A Theory of Human Image Understanding. Psychol Rev.

Biederman I, Gerhardstein PC (1993) Recognizing Depth-Rotated Objects: Evidence and Conditions for Three-Dimensional Viewpoint Invariance. J Exp Psychol Hum Percept Perform.

Bird A, Tobin E (2018) Natural Kinds. In: The Stanford Encyclopedia of Philosophy, Spring 201. (Zalta EN, ed). Metaphysics Research Lab, Stanford University.

Boedecker J, Obst O, Lizier JT, Mayer NM, Asada M (2012) Information processing in echo state networks at the edge of chaos. Theory Biosci 131:205–213.

Bosman CA, Schoffelen JM, Brunet N, Oostenveld R, Bastos AM, Womelsdorf T, Rubehn B, Stieglitz T, De Weerd P, Fries P (2012) Attentional Stimulus Selection through Selective Synchronization between Monkey Visual Areas. Neuron 75:875–888 Available at: http://dx.doi.org/10.1016/j.neuron.2012.06.037.

Brigard F De (2012) The role of attention in conscious recollection. Front Psychol 3:1–10.

Briggs F, Mangun GR, Usrey WM (2013) Attention enhances synaptic efficacy and the signal-to-noise ratio in neural circuits. Nature 499:476–480 Available at: http://www.nature.com/doifinder/10.1038/nature12276.

Brodeur, Mathieu B.; Guerard, Katherine; Bouras M (2014) Bank of Standardized Stimuli (BOSS) Phase II : 930 New Normative Photos. PLOS 9.

Burge J, Geisler WS (2011) Optimal defocus estimation in individual natural images. Proc Natl Acad Sci U S A 108:16849–16854 Available at: https://www.ncbi.nlm.nih.gov/pubmed/21930897.

Buschman TJ, Kastner S (2016) From behavior to neural dynamics: An integrated theory of attention. 88:127–144.

Buzsáki G (2020) The brain–cognitive behavior problem: A retrospective. eNeuro.

Buzsaki G, Draguhn A (2004) Neuronal oscillations in cortical networks. Science (80-) 304:1926–1929 Available at: https://www.ncbi.nlm.nih.gov/pubmed/15218136.

Buzsáki G, Logothetis N, Singer W (2013) Scaling brain size, keeping timing: Evolutionary preservation of brain rhythms. Neuron.

Buzśaki G, Wang XJ (2012) Mechanisms of gamma oscillations. Annu Rev Neurosci.

Capotosto P, Babiloni C, Romani GL, Corbetta M, San C, Cassino R (2009) Frontoparietal Cortex Controls Spatial Attention through Modulation of Anticipatory Alpha Rhythms. J Neurosci 29:5863–5872.

Capotosto P, Baldassarre A, Sestieri C, Spadone S, Romani GL, Corbetta M (2017) Task and Regions Specific Top-Down Modulation of Alpha Rhythms in Parietal Cortex. Cereb Cortex 27:4815–4822.

Carlson T, Tovar DA, Alink A, Kriegeskorte N (2013) Representational dynamics of object vision: The first 1000 ms. J Vis.

Carlson TA, Brendan Ritchie J, Kriegeskorte N, Durvasula S, Ma J (2014) Reaction time for object categorization is predicted by representational distance. J Cogn Neurosci.

Carrasco M, Barbot A (2019) Spatial attention alters visual appearance. Curr Opin Psychol 29:56–64 Available at: https://doi.org/10.1016/j.copsyc.2018.10.010.

Chalmers DJ (2018) Facing Up to the Problem of Consciousness. In: Consciousness and Emotion in Cognitive Science.

Chun MM, Golomb JD, Turk-Browne NB (2011) A taxonomy of external and internal attention. Annu Rev Psychol 62:73–101 Available at: https://www.ncbi.nlm.nih.gov/pubmed/19575619.

Cichy RM, Pantazis D, Oliva A (2014) Resolving human object recognition in space and time. Nat Neurosci.

Cohen EH, Tong F (2015) Neural Mechanisms of Object-Based Attention. Cereb Cortex:1080–1092.

Cohen MA, Konkle T, Rhee JY, Nakayama K, Alvarez GA (2014) Processing multiple visual objects is limited by overlap in neural channels. Proc Natl Acad Sci U S A.

Cohen MA, Nakayama K, Konkle T, Stantić M, Alvarez GA (2015) Visual awareness is limited by the representational architecture of the visual system. J Cogn Neurosci.

Corbetta M, Kincade JM, Ollinger JM, McAvoy MP, Shulman GL (2000) Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nat Neurosci 3:292–297.

Corbetta M, Miezin F, Dobmeyer S, Shulman G, Petersen S (1990) Attentional modulation of neural processing of shape, color, and velocity in humans. Science (80-) 248:1556–1559.

Corbetta M, Patel G, Shulman GL (2008) The Reorienting System of the Human Brain: From Environment to Theory of Mind. Neuron 58:306–324.

Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 3:201–215 Available at: https://www.ncbi.nlm.nih.gov/pubmed/11994752.

Coultrip R, Granger R, Lynch G (1992) A cortical model of winner-take-all competition via lateral inhibition. Neural Networks.

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21 Available at: https://www.ncbi.nlm.nih.gov/pubmed/15102499.

Desimone R, Duncan J (1995) Neural Mechanisms of Selective Visual Attention. Annu Rev Neurosci 18:193–222.

DiQuattro NE, Geng JJ (2011) Contextual knowledge configures attentional control networks. J Neurosci.

Duncan J, Humphreys GW (1989) Visual Search and Stimulus Similarity. Psychol Rev.

Eason RG (1981) Visual evoked potential correlates of early neural filtering during selective attention. Bull Psychon Soc 18:203–206.

Egly R, Driver J, Rafal RD (1994) Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. J Exp Psychol Gen 123:161–177 Available at: https://www.ncbi.nlm.nih.gov/pubmed/8014611.

Eimer M (1996) ERP modulations indicate the selective processing of visual stimuli as a result of transient and sustained spatial attention. Psychophyiology 33:13–21.

Epstein R, Harris A, Stanley D, Kanwisher N (1999) The Parahippocampal Place Area: Recognition, Navigation, or Encoding? Neuron 23:115–125.

Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. Nature 392:6–9.

Fedorenko E, Duncan J, Kanwisher N (2013) Broad domain generality in focal regions of frontal and parietal cortex. Proc Natl Acad Sci U S A.

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex.

Fries P (2005) A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. Trends Cogn Sci 9:474–480.

Fries P, Nikolić D, Singer W (2007) The gamma cycle. Trends Neurosci 30:309–316.

Fries P, Reynolds JH, Rorie AE, Desimone R (2001) Modulation of Oscillatory Neuronal Synchronization by Selective Visual Attention. Science (80-) 291:1560–1564.

Gauthier I, Tarr MJ (2016) Visual Object Recognition: Do We (Finally) Know More Now Than We Did? Annu Rev Vis Sci.

Gauthler I, Behrmann M, Tarr MJ (1999) Can face recognition really be dissociated from object recognition? J Cogn Neurosci 11:349–370.

Geisler WS, Perry JS (2011) Statistics for optimal point prediction in natural images. J Vis 11:14 Available at: https://www.ncbi.nlm.nih.gov/pubmed/22011382.

Geng JJ, Mangun GR (2011) Right temporoparietal junction activation by a salient contextual cue facilitates target discrimination. Neuroimage.

Geng JJ, Vossel S (2013) Re-evaluating the role of TPJ in attentional control: Contextual updating? Neurosci Biobehav Rev.

Giesbrecht B, Woldorff MG, Song AW, Mangun GR (2003) Neural mechanisms of top-down control during spatial and feature attention. Neuroimage 19:496–512 Available at: https://www.ncbi.nlm.nih.gov/pubmed/12880783.

Green JJ, Doesburg SM, Ward LM, McDonald JJ (2011) Electrical neuroimaging of voluntary audiospatial attention: Evidence for a supramodal attention control network. J Neurosci.

Greenberg AS, Esterman M, Wilson D, Serences JT, Yantis S (2010) Control of spatial and feature-based attention in frontoparietal cortex. J Neurosci.

Grill-Spector K (2003) The neural basis of object perception. Curr Opin Neurobiol.

Grill-Spector K, Kanwisher N (2005) Visual recognition: As soon as you know it is there, you know what it is. Psychol Sci.

Grill-Spector K, Malach R (2004) THE HUMAN VISUAL CORTEX. Annu Rev Neurosci.

Grill-Spector K, Weiner KS (2014) The functional architecture of the ventral temporal cortex and its role in categorization. Nat Rev Neurosci.

Haxby J V., Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science (80-).

Heinze HJ, Mangun GR, Burchert W, Hinrichs H, Scholz M, Münte TF, Gos A, Johannes MSS, Hundeshagen H, Gazzaniga MS, Hillyard SA (1994) Combined spatial and temporal imaging of brain activity during visual selective attention in humans. Nature 372:543–546.

Henriksson L, Mur M, Kriegeskorte N (2019) Rapid Invariant Encoding of Scene Layout in Human OPA. Neuron 103:1–11 Available at: https://doi.org/10.1016/j.neuron.2019.04.014.

Hillyard SA, Vogel EK, Luck SJ (1998) Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. Philos Trans R Soc B Biol Sci.

Hommel B, Chapman CS, Cisek P, Neyedli HF, Song J-H, Welsh TN (2019) No one knows what attention is. Attention, Perception, Psychophys:2288–2303.

Hopfinger JB, Buonocore MH, Mangun GR (2000) The neural mechanisms of top-down attentional control. Nat Neurosci 3:284–291 Available at: https://www.ncbi.nlm.nih.gov/pubmed/10700262.

Jensen O, Kaiser J, Lachaux JP (2007) Human gamma-frequency oscillations associated with attention and memory. Trends Neurosci.

Jensen O, Mazaheri A (2010) Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front Hum Neurosci 4:186 Available at: https://www.ncbi.nlm.nih.gov/pubmed/21119777.

Jurcak V, Tsuzuki D, Dan I (2007) 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. Neuroimage 34:1600–1611.

Kanwisher N, Mcdermott J, Chun MM (1997) The Fusiform Face Area : A Module in Human Extrastriate Cortex Specialized for Face Perception. J Neurosci 17:4302–4311.

Kanwisher N, Yovel G (2006) The fusiform face area: a cortical region specialized for the perception of faces. Philos Trans R Soc B Biol Sci 361:2109–2128.

Kastner S, De Weerd P, Desimone R, Ungerleider LG (1998) Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. Science (80-).

Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG (1999) Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron 22:751–761.

Kersey AJ, Clark TS, Lussier CA, Mahon BZ, Cantlon JF (2016) Development of Tool Representations in the Dorsal and Ventral Visual Object Processing Pathways. Cereb Cortex 26:3135–3145 Available at: https://www.ncbi.nlm.nih.gov/pubmed/26108614.

Khaligh-Razavi SM, Kriegeskorte N (2014) Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. PLoS Comput Biol.

Klimesch W (2012) Alpha-based oscillations, attention, and controlled access to stored information. Trends Cogn Sci 16:606–617 Available at: http://dx.doi.org/10.1016/j.tics.2012.10.007.

Klimesch W, Sauseng P, Hanslmayr S (2007) EEG alpha oscillations: The inhibition-timing hypothesis. Brain Res Rev 53:63–88.

Konen CS, Kastner S (2008) Two hierarchically organized neural systems for object information in human visual cortex. Nat Neurosci.

Konkle T, Caramazza A (2013) Tripartite organization of the ventral stream by animacy and object size. J Neurosci.

Konkle T, Oliva A (2012) A Real-World Size Organization of Object Responses in Occipitotemporal Cortex. Neuron.

Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M (2013) The ventral visual pathway: An expanded neural framework for the processing of object quality. Trends Cogn Sci.

Kriegeskorte N (2015) Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. Annu Rev Vis Sci.

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008) Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. Neuron.

Langton CG (1990) Computation at the edge of chaos: Phase transitions and emergent computation. Phys D Nonlinear Phenom 42:12–37.

Lehky SR, Tanaka K (2016) Neural representation for object recognition in inferotemporal cortex. Curr Opin Neurobiol.

Liu T, Hou Y (2013) A hierarchy of attentional priority signals in human frontoparietal cortex. J Neurosci 33:16606–16616 Available at: https://www.ncbi.nlm.nih.gov/pubmed/24133264.

Liu Y, Bengson J, Huang H, Mangun GR, Ding M (2016) Top-down Modulation of Neural Activity in Anticipatory Visual Attention: Control Mechanisms Revealed by Simultaneous EEG-fMRI. Cereb Cortex 26:517–529.

Logothetis NK, Wandell BA (2004) Interpreting the BOLD Signal. Annu Rev Physiol.

Long B, Yu CP, Konkle T (2018) Mid-level visual features underlie the high-level categorical organization of the ventral stream. Proc Natl Acad Sci U S A.

Luck SJ (Steven J (2014) An Introduction to the Event-Related Potential Technique, second edition.

Luck SJ, Chelazzi L, Hillyard SA, Desimone R (1997) Neural Mechanisms of Spatial Selective Attention in Areas V1 , V2 , and V4 of Macaque Visual Cortex. J

Neurophysiol 77:24–42.

Luck SJ, Gold JM (2008) The Construct of Attention in Schizophrenia. Biol Psychiatry.

Luck SJ, Hillyard SA, Mouloua M, Woldorff MG, Clark VP, Hawkins HL (1994) Effects of Spatial Cuing on Luminance Detectability: Psychophysical and Electrophysiological Evidence for Early Selection. J Exp Psychol Hum Percept Perform.

Luck SJ, Vecera SP (2002) Chapter 6 : Attention. In: Stevens's Handbook of Experimental Psychology: Vol.1. Sensation and Perception.

Luck SJ, Woodman GF, Vogel EK (2000) Event-related potential studies of attention. Trends Cogn Sci 4:432–440.

Ma DS, Correll J, Wittenbrink B (2015) The Chicago face database: A free stimulus set of faces and norming data. Behav Res Methods.

Malcolm GL, Shomstein S (2015) Object-based attention in real-world scenes. J Exp Psychol Gen.

Mangun GR (1995) Neural mechanisms of visual selective attention. Psychophysiology.

Mangun GR, Buonocore MH, Girelli M, Jha AP (1998) ERP and fMRI Measures of Visual Spatial Selective Attention. Hum Brain Mapp 6:383–389.

Mangun GR, Hillyard SA (1991) Modulations of Sensory-Evoked Brain Potentials Indicate Changes in Perceptual Processing During Visual-Spatial Priming. J Exp Psychol Hum Percept Perform 17:1057–1074.

Marr D (1982) Vision: a computational investigation into the human representation and processing of visual information. Vis a Comput Investig into Hum Represent Process Vis information.

Martínez A, Anllo-Vento L, Sereno MI, Frank LR, Buxton RB, Dubowitz DJ, Wong EC, Hinrichs H, Heinze HJ, Hillyard SA (1999) Involvement of striate and extrastriate visual cortical areas in spatial attention. Nat Neurosci 2:364–369.

Mazaheri A, Jensen O (2008) Asymmetric amplitude modulations of brain oscillations generate slow evoked responses. J Neurosci.

Mervis CB, Rosch E (1981) Categorization of Natural Objects. Annu Rev Psychol.

Mirollo RE, Strogatz SH (1990) Synchronization of pulse-coupled biological oscillators. SIAM J Appl Math.

Mitchell JF, Sundberg KA, Reynolds JH (2009) Article Spatial Attention Decorrelates Intrinsic Activity Fluctuations in Macaque Area V4 SHUFFLE. Neuron 63:879–888 Available at: http://dx.doi.org/10.1016/j.neuron.2009.09.013.

Moore CM, Yantis S, Vaughan B (1998) Object-based visual selection: Evidence from Perceptual Completion. Psychol Sci.

Moore T, Zirnsak M (2017) Neural Mechanisms of Selective Visual Attention. Annu Rev Psychol 68:47–72.

Moran J, Desimone R (1985) Selective Attention Gates Visual Processing in the Extrastriate Cortex. Science (80-) 229:782–784.

Mruczek REB, von Loga IS, Kastner S (2013) The representation of tool and non-tool object information in the human intraparietal sulcus. J Neurophysiol.

Mur M, Ruff DA, Bodurka J, De Weerd P, Bandettini PA, Kriegeskorte N (2012) Categorical, yet graded-single-image activation profiles of human category-selective cortical regions. J Neurosci.

Noah S, Mangun GR (2020) Recent evidence that attention is necessary, but not sufficient, for conscious perception. Ann N Y Acad Sci 1464:52–63.

Noah S, Powell T, Khodayari N, Olivan D, Ding M, Mangun GR (2020) Neural Mechanisms of Attentional Control for Objects: Decoding EEG Alpha When Anticipating Faces, Scenes, and Tools. J Neurosci.

Nobre AC, Sebestyen GN, Gitelman DR, Mesulam MM, Frackowiak RSJ, Frith CD (1997) Functional localization of the system for visuospatial attention using positron emission tomography. Brain 120:515–533.

O'Connor DH, Fukui MM, Pinsk MA, Kastner S (2002) Attention modulates responses in the human lateral geniculate nucleus. Nat Neurosci 5:1203–1209 Available at: https://www.ncbi.nlm.nih.gov/pubmed/12379861.

O'Craven KM, Downing PE, Kanwisher N (1999) fMRI evidence for objects as the units of attentional selection. Nature 401:584–587 Available at: https://www.ncbi.nlm.nih.gov/pubmed/10524624.

Palmeri TJ, Gauthier I (2004) Visual object understanding. Nat Rev Neurosci.

Peelen M V., Fei-Fei L, Kastner S (2009) Neural mechanisms of rapid natural scene categorization in human visual cortex. Nature.

Peelen M V., Kastner S (2011) A neural basis for real-world visual search in human occipitotemporal cortex. Proc Natl Acad Sci U S A.

Peelen M V, Kastner S (2014) Attention in the real world: toward understanding its neural basis. Trends Cogn Sci 18:242–250 Available at: https://www.ncbi.nlm.nih.gov/pubmed/24630872.

Perry CJ, Fallah M (2014) Feature integration and object representations along the dorsal stream visual hierarchy. Front Comput Neurosci.

Petersen SE, Posner MI (2012) The Attention System of the Human Brain : 20 Years After. Annu Rev Psychol 35:73–89.

Pfurtscheller G, Stancák A, Neuper C (1996) Event-related synchronization (ERS) in the alpha band - an electrophysiological correlate of cortical idling: A review. Int J Physiol 24:1–2.

Popov T, Kastner S, Jensen O (2017) FEF-controlled alpha delay activity precedes stimulus-induced gamma-band activity in visual cortex. J Neurosci 37:4117–4127.

Posner MI (1980) Orienting of attention. Q J Exp Psychol 32:3–25 Available at: https://www.ncbi.nlm.nih.gov/pubmed/7367577.

Posner MI, Snyder CR, Davidson BJ (1980) Attention and the detection of signals. J Exp Psychol 109:160–174 Available at: https://www.ncbi.nlm.nih.gov/pubmed/7381367.

Rajagovindan R, Ding M (2011) From Prestimulus Alpha Oscillation to Visual-evoked Response: An Inverted-U Function and Its Attentional Modulation Rajasimhan Rajagovindan. J Cogn Neurosci 23:1379–1394.

Rajalingham R, Issa EB, Bashivan P, Kar K, Schmidt K, DiCarlo JJ (2018) Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. J Neurosci.

Righi G, Peissig JJ, Tarr MJ, Righi G, Peissig JJ, Tarr MJ, Righi G, Peissig JJ, Tarr MJ (2012) Recognizing disguised faces. Vis cogn 20.

Robinson TM (1987) Heraclitus: Fragments. University of Toronto Press.

Romei V, Brodbeck V, Michel C, Amedi A, Pascual-Leone A, Thut G (2008) Spontaneous fluctuations in posterior alpha-band EEG activity reflect variability in excitability of human visual areas. Cereb Cortex 18:2010–2018 Available at: https://www.ncbi.nlm.nih.gov/pubmed/18093905.

Rovelli C (2018) The Order of Time. Penguin Books Limited. Available at: https://books.google.com/books?id=4ow7DwAAQBAJ.

Saalmann YB, Pinsk MA, Wang L, Li X, Kastner S (2012) The pulvinar regulates information transmission between cortical areas based on attention demands. Science (80-) 337:753–756 Available at: https://www.ncbi.nlm.nih.gov/pubmed/22879517.

Salmela V, Salo E, Salmi J, Alho K (2018) Spatiotemporal Dynamics of Attention Networks Revealed by Representational Similarity Analysis of EEG and fMRI. Cereb Cortex.

Sauseng P, Klimesch W, Stadler W, Schabus M, Doppelmayr M, Hanslmayr S, Gruber WR, Birbaumer N (2005) A shift of visual spatial attention is selectively associated with human EEG alpha activity. Eur J Neurosci 22:2917–2926.

Scholl BJ (2001) Objects and attention: the state of the art. Cognition 80:1–46 Available at: https://www.ncbi.nlm.nih.gov/pubmed/11245838.

Scolari M, Seidl-Rathkopf KN, Kastner S (2015) Functions of the human frontoparietal attention network: Evidence from neuroimaging. Curr Opin Behav Sci 1:32–39 Available at: https://www.ncbi.nlm.nih.gov/pubmed/27398396.

Serences JT, Schwarzbach J, Courtney SM, Golay X, Yantis S (2004) Control of object-based attention in human cortex. Cereb Cortex.

Shomstein S, Yantis S (2004) Control of attention shifts between vision and audition in human cortex. J Neurosci.

Silver MA, Kastner S (2009) Topographic maps in human frontal and parietal cortex. Trends Cogn Sci.

Simoncelli EP, Olshausen BA (2001) Natural Image Statistics and Neural Representation. Annu Rev Neurosci.

Snyder AC, Foxe JJ (2010a) Anticipatory Attentional Suppression of Visual Features Indexed by Oscillatory Alpha-Band Power Increases:A High-Density Electrical Mapping Study. J Neurosci 30:4024–4032 Available at: http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.5684-09.2010.

Snyder AC, Foxe JJ (2010b) Anticipatory attentional suppression of visual features indexed by oscillatory alpha-band power increases: a high-density electrical mapping study. J Neurosci 30:4024–4032 Available at: https://www.ncbi.nlm.nih.gov/pubmed/20237273.

Spagna A, Mackie MA, Fan J (2015) Supramodal executive control of attention. Front Psychol.

Spoerer CJ, McClure P, Kriegeskorte N (2017) Recurrent convolutional neural networks: A better model of biological object recognition. Front Psychol.

Szczepanski SM, Pinsk MA, Douglas MM, Kastner S, Saalmann YB (2013) Functional and structural architecture of the human dorsal frontoparietal attention network.

Proc Natl Acad Sci U S A.

Tanaka K (1997) Mechanisms of visual object recognition: Monkey and human studies. Curr Opin Neurobiol.

Tarr MJ, Bülthoff HH (1998) Image-based object recognition in man, monkey and machine. Cognition.

Tarr MJ, Gauthier I (2000) FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. Nat Neurosci.

Thut G, Nietzel A, Brandt SA, Pascual-leone A (2006) Alpha-Band Electroencephalographic Activity over Occipital Cortex Indexes Visuospatial Attention Bias and Predicts Visual Target Detection. J Neurosci 26:9494–9502.

Tipper SP, Behrmann M (1996) Object-Centered Not Scene-Based Visual Neglect. J Exp Psychol Hum Percept Perform 22:1261–1278.

Tong F (2018) Foundations of Vision. In: Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience.

Tootell RBH, Hadjikhani N, Hall EK, Marrett S, Vanduffel W, Vaughan JT, Dale AM (1998) The Retinotopy of Visual Spatial Attention. Neuron 21:1409–1422.

Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cogn Psychol 12:97–136 Available at: https://www.ncbi.nlm.nih.gov/pubmed/7351125.

Van Dijk H, Van Der Werf J, Mazaheri A, Medendorp WP, Jensen O (2010) Modulations in oscillatory activity with amplitude asymmetry can produce cognitively relevant event-related responses. Proc Natl Acad Sci U S A.

Van Voorhis S, Hillyard SA (1977) Visual evoked potentials and selective attention to points in space. Percept Psychophys 22:54–62.

Veniero D, Gross J, Morand S, Duecker F, Sack AT, Thut G (2021) Top-down control of visual cortex by the frontal eye fields through oscillatory realignment. Nat Commun 12 Available at: http://dx.doi.org/10.1038/s41467-021-21979-7.

Vossel S, Geng JJ, Fink GR (2014) Dorsal and ventral attention systems: Distinct neural circuits but collaborative roles. Neuroscientist.

Wang W, Viswanathan S, Lee T, Grafton ST (2016) Coupling between theta oscillations and cognitive control network during cross-modal visual and auditory attention: Supramodal vs modality-specific mechanisms. PLoS One.

Wojciulik E, Kanwisher N (1999) The generality of parietal involvement in visual attention. Neuron.

Woodman GF, Arita JT, Luck SJ (2009) A cuing study of the N2pc component: An index of attentional deployment to objects rather than spatial locations. Brain Res 1297:101–111 Available at: http://dx.doi.org/10.1016/j.brainres.2009.08.011.

Worden MS, Foxe JJ, Wang N, Simpson G V (2000) Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. J Neurosci 20:RC63 Available at: https://www.ncbi.nlm.nih.gov/pubmed/10704517.

Zumer JM, Scheeringa R, Schoffelen JM, Norris DG, Jensen O (2014) Occipital Alpha Activity during Stimulus Processing Gates the Information Flow to Object-Selective Cortex. PLoS Biol 12.