Spring 2018

# Towards Improving Learning with Consumer-Grade, Closed-Loop, Electroencephalographic Neurofeedback

Zall Soren Hirschstein
*Bard College*, zh8605@bard.edu

## Recommended Citation

Hirschstein, Zall Soren, "Towards Improving Learning with Consumer-Grade, Closed-Loop, Electroencephalographic Neurofeedback" (2018). *Senior Projects Spring 2018*. 206.
https://digitalcommons.bard.edu/senproj_s2018/206

# Towards Improving Learning with Consumer-Grade, Closed-Loop, Electroencephalographic Neurofeedback

Senior Project Submitted to
The Division of Science, Mathematics, and Computing
of Bard College

by

Zall Hirschstein

Annandale-on-Hudson, New York
May, 2018

## Table of Contents

**ABSTRACT**

**INTRODUCTION**

**Section 1: Learning**

**Section 2: Technology**

**Section 3: EEG**

**METHODS**

## ABSTRACT

Learning is an enigmatic process composed of a multitude of cognitive systems that are functionally and neuroanatomically distinct. Nevertheless, two undeniable pillars which underpin learning are attention and memory; to learn, one must attend, and maintain a representation of, an event. Psychological and neuroscientific technologies that permit researchers to “mind-read” have revealed much about the dynamics of these distinct processes that contribute to learning. This investigation first outlines the cognitive pillars which support learning and the technologies that permit such an understanding. It then employs a novel task—the amSMART paradigm—with the goal of building a real-time, closed-loop, electroencephalographic (EEG) neurofeedback paradigm using consumer-grade brain-computer interface (BCI) hardware. Data are presented which indicate the current status of consumer-grade BCI for EEG cognition classification and enhancement, and directions are suggested for the developing world of consumer neurofeedback.

# INTRODUCTION

## Section 1: Learning

### *1.1 What* is *Learning?*

The only sound in the library at this time of night is the low hum that fluorescent bulbs emit when they flicker. Glowing rods sway overhead like pendulums. The oversized clock hung on exposed brick ticks 3:00 AM. You have seven hours to see how much meaning can float off the pages of notes in front of your eyes and into the organ behind them.

If an outsider were to observe this situation, they'd likely say that you'd be better off sacrificing this late-night quest to shove a few facts up your sleeve for the benefits of a decent, if half-, night of sleep. But sitting alone, at the end of a long table, papers piled high around you and the blue-tinted fluorescent lights dissuading your brain from understanding that the sun set over nine hours ago, there's no one—or nothing—to tell you this.

Learning is a dynamic and enigmatic process. In the above scenario, the best recourse for your learning is likely *not* trying to learn anything more that night. How can this be?

Even attempting to define learning—which seems like an important foundation for any investigation into its nature—is an elusive project; or at the very least, contentious: There exist innumerable "types" of learning, each distinguished from the others by the theoretical framework of which group of theorists attempt to define it. Under more traditional, psychological frameworks, learning might be understood chiefly as the acquisition of knowledge and skills (Shuell, 1986), but in the current era, the concept could cover a much larger field of sub-concepts that embrace an interacting body of emotional and social dimensions—or even a dimension for how tired the subject is—for example (Illeris, 2009). Accordingly, it is difficult to identify from where one might derive any foundations for an investigation into learning—where to find the solid ground from which to leap.

A loose description might be enough to move forward: Learning is “any process that in a living organisms leads to permanent capacity change and which is not solely due to biological maturation or ageing” (Illeris, 2007, p. 3). Thus, in order to say that something can learn, it must update a response to an environmental factor based on previous encounters with environmental factors.

Following this line of thought, learning seems to involves two fundamental things: (1) The separation of “pertinent” from “not pertinent” information, and (2) the maintenance of that information. That is, an organism must attend a stimulus and its effect and then employ a predictive representation of that relationship in the future; an organism must notice something and be changed by it. While not the sole supporters of learning processes, attention and memory seem to provide much of the cognitive infrastructure. Like learning, however, there are many heads bobbing underneath the singular umbrellas of “attention” and “memory.”

This paper traces the roots of how attention and memory have become defined, surveys the technological tools used to reveal these definitions, and investigates a novel means of attempting to improve learning by combining the psychological frameworks outlined herein with advancements in neuroscientific technologies.

### *1.2 Attention*

The brain has plenty of remarkable feats to accomplish, but a chief purpose of its nature is to navigate the perceived world—to find a route; semblance of signal out of sensory noise. It seems logical then that the capacity to remain focused on a feature of the environment or on a singular task is a vital underpinning of many cognitive functions, including learning.

Phenomena such as inattentional blindness (Simons & Chabris, 1999)—a failure of the attentional system in which conspicuous material is not noticed due to the presence of distractors—demonstrate the continual necessity of distinguishing pertinent from extraneous information in sensory data. And such failures of attention can be dangerous: Indeed, for example, attentional deficits predict rates of car crashes and treatment with first line pharmaceuticals for ADHD (attention-deficit hyperactivity disorder) has been shown to reduce these and other problematic driving outcomes (Jerome, Segal, & Habinski, 2006; Barkley, Murphy, O'Connell, & Connor, 2005). Elucidating and understanding the inner workings of attention therefore remains a core aim of subfields in cognitive science, both for those attempting to understand its dynamics and basis (e.g., Heinke & Humphreys, 2005) as well as how to reduce the rates and effects of its failures (e.g., Romer, Lee, McDonald, & Winston, 2014).

Selection is obviously a crucial component of attention because it is more or less entailed by the concept; in the act of attending something, its characteristics are selected from less relevant and (thereby) distracting information (Remington & Folk, 2001). Imagine being at a boisterous work event. Cocktails are being served and your boss is being particularly loud (no causal relationship is being assumed, but there is no doubt that one might exist); being so loud, in fact, that—despite the fact that she is at the other side of the room—you can't even hear the one co-worker next to you with whom you are trying to have a conversation! Your ears may be receiving all the information necessary to parse out your co-worker's words, but the signal is all mixed up with the noise of your boss's shrill laugh. The brain's job here is to process the information from the ears such that the relevant signal (the words of your co-worker) is distinguishable. Maybe the best feature to focus on in order to "select" your coworker's words would be the pitch of their voice (because it's much higher or lower than that of your boss); maybe you're facing away from your boss, but towards your coworker and can use spatial inference to select the information stream in "front" of your ears. The phenomenon's name comes from this common example: the Cocktail Party Effect (Cherry, 1953). To select the signal, the brain must segregate and rummage through different components of the audio information [i.e., spatial, spectral (frequency/amplitude), intensity, content, etc.]—which are categorized via perceptual segregation and perceptual grouping (e.g., temporal grouping, spectral grouping, etc.; Bregman, 1990; Treisman, 1982). For ease of understanding, a visual analog of perceptual grouping is seen in Figure 1.

Much of the early work on attention was also inspired by studies that demonstrated a robust effect of spatial orienting, accounting for data showing attentional bias towards areas in visual space previously occupied by other stimuli (or auditory "space" (Smith et al., 2010); and the interaction of these two via multisensory integration (Nardo, Santangelo, & Macaluso, 2014; Razavi, O'Neill, & Paige, 2007)), referred to as spatially predictive cues (Posner, 1980; Chica, Martín-Arévalo, Botta, & Lupiáñez, 2014). Thus, the presence of a potentially predictive cue like the negative "square" in Figure 1 might bias visual attention in future trials to that location in the search space (Posner, 1980).

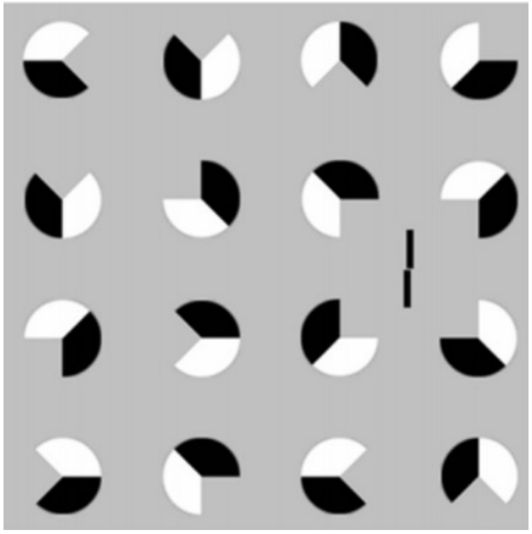


**Figure 1.** Despite the fact that no rectangle is actually present in this image, the four collinear "pacmen" in the middle right of the figure (with the two rays inside) demonstrate the brain's ability and propensity to select potentially pertinent signals (here, simple patterns—a square) in noisy information from perception and group them. Adapted from "Perceptual Organization, Visual Attention, and Objecthood" by Kimchi, Yeshurun, Spehar, & Pirkner, 2015, *Vision Research*, 126, 34-51.

While no one doubts the inherency of selection in mental processing, within this conceptual scheme of attention-as-selection, one can imagine that a central question is: How and when does this selection occur? While the exact mechanisms and dynamics of this system are far from being fully understood, one of the most long-standing and contentious debates regarding the attentional system orients around the temporal characteristic of distractor influence and the resulting mechanistic implications of either an “early” selection or a “late” selection effect of attention.

The early selection account—most prominently spearheaded by Broadbent (1958) and Treisman (1969; though their accounts do have some major differences related to the nature of the attentional filter)—maintains that perception is a capacity-limited process requiring selection before passing along the information to the rest of the mind. Thus this selection occurs “early”—only after basic feature processing has offered enough information for selection to occur—and unattended stimuli are not passed on to the mechanisms of conscious perception (in full for Treisman, 1969; Lavie, 1995). Data supporting this early selection theory were mostly derived from dichotic listening experiments in which participants attended two audio streams being presented to either ear. The participants were asked to attend audio on the basis of a feature, such as the perceived gender of the speaker or which ear the audio was being sent into (Broadbent, 1958). These experiments typically relied on direct measures of awareness (e.g., “shadowing”—repeating the words of—the attended or unattended person’s voice). But while these direct measures often indicated that there might be an early bottleneck for perception, more indirect measures suggested that such a bottleneck was not always present (Murphy, Groeger, & Greene, 2016).

The “opposing” opinion in the literature is the “late” selection account. This theory was advanced most prominently by Deutch and Deutch (1963), who proposed that all stimuli are processed indiscriminately and input to the perceptual system to be subsequently selected for response- and memory-making (Treisman & Geffen, 1967). Thus, perception would not be capacity-limited but an *un*limited filter applied to all sensory input. By the early 1990s, while some evidence had suggested that the early selection account is not a holistic view of the nature of the attentional filtering, likewise the late selection hypothesis did not account for all of the findings in the field, either. How can these conflicts of data be compatible?

In the 1990's, a newer, more nuanced formulation of attention began brewing. Yantis and Johnston (1990) proposed a hybrid model, suggesting that the location of the selective filter might shift location in the processing pipeline based on what kind of processing was going on (i.e., how much and what kind of sensory information). Building on this, Lavie & Tsal (1994) coined a new account, "load theory," which promoted an intermediary solution between early and late selection theories (cf. Benoni & Tsal, 2013). Evidence supporting this account suggests that the degree to which a person's perceptual systems are loaded with task-relevant information greatly affects how the attention systems performs its selection. That is, perceptual load, as a function of the number of items or groups of items contained in the search space to be processed (set-size; Lavie, 1995), determines the efficiency and efficacy of distractor rejection. This theory has grown in popularity since its inception, at least in part, due to the fact that it formulated a middle-ground between early and late selection theories in which the success or failure of attentional selection depends on the current demands of the system (Murphy et al., 2016).

As attention is so inherent to experience, it's clear why there is so much debate about its characteristics. But a unified theory of attention that applies to both the "voluntary" and "involuntary" instances is difficult: Attention is not a unitary system at a location and time in the processing pipeline of the brain; in fact, there are two independent, generally identified pathways of attention: Bottom-up and top-down—these are also referred to as *stimulus-driven* and *goal-oriented* attention respectively (Corbetta & Shulman, 2002; Jonides, 1981). Think of it this way: You're sitting in your living room listening intently to a Mendelssohn piano concerto—what you might consider the most beautiful music ever composed. But nonetheless, when the smoke alarm goes off in the kitchen (because, in the midst of your audial reverie, you've long neglected the tea kettle you put on the stove 15 minutes ago) you quickly "forget" about—more properly, your attention is drawn away from—the music and you rush into the kitchen to open a window and turn off the kettle. This would be a case of your bottom-up (stimulus-driven) attention system taking precedence over your top-down (goal-oriented) attention, drawing your mind away from the music and towards the most salient, threatening stimulus in the environment. Novel, unexpected, and potentially dangerous stimuli in the environment drive this bottom-up attention system while factors such as goals, knowledge, and expectations construct top-down attention. It seems necessary that this bottom-up system be the circuit-breaker for the top-down one; if humans were so prone to being carried away by music that they failed to recognize the salient growl of a predator behind them, it would be doubtful that they would have gotten far at all. It's proposed that the interaction of these two independent attentional processes dictate when and to what one pays attention (Corbetta & Shulman, 2002).

But this theory is also inadequate in explaining all the dynamics of attention. For example, Awh, Belopolsky, & Theeuwes (2012) put forward a model of attention that goes beyond this dichotomy of top-down vs. bottom-up attention. Their reconceptualization aligns with data that evidences a lingering bias in selection for features previously selected ("selection history") and for those that people have been previously rewarded for selecting ("reward history")—features that the authors claim are insufficiently accounted for by the dichotomous attention model. Thus, they suggest expanding on a framework by which people create priority maps—conceptual representations that integrate multiple selection influences, such as salience or position in the visual space—to include these other elements (e.g., selection and reward history; Awh, Belopolsky, & Theeuwes, 2012; Klink, Jentgens, & Lorteije, 2014). So how do these researchers build and test these models?

The mode of attention most easily evident for humans is that of visual attention (as was demonstrated in using Figure 1 as an aid for understanding). Accordingly, one can imagine that attentional research is often visual in nature; thus, a prominent tool for investigating attentional processes is the visual search task. Everyone is, in fact, very familiar with this task because everyone performs it all the time: Visual search refers to a perceptual process that requires attention in which one actively scans the visual environments for a particular object or feature (called the target) among other objects and features (called distractors; Treisman & Gelade, 1980). So when someone named Sam is looking for their name in a pile of nametags at a conference, they might determine the letter S to be an important object or three-letter names to be a feature of interest—ideally, both at the same time. But also a possibility is that while searching for these objects and features, they might spot the bright red in the logo of the institution they represent and their eye might be instantly drawn to it—voila!—Sam has found their nametag. The former means of finding the nametag in this situation would recruit top-down processes (based on goals—finding a specified letter or number of letters); the latter, bottom-up processes, based on sensory salience—here, the color of the company logo (but if Sam were actively searching for that color, then that would represent a top-down process which would help that color feature “pop” out; more on this soon).

In the laboratory, the visual search tasks that are used are usually much simpler than this complex problem for the sake of not having to control so many variables, and may look like the search spaces in Figures 2a and 2b.

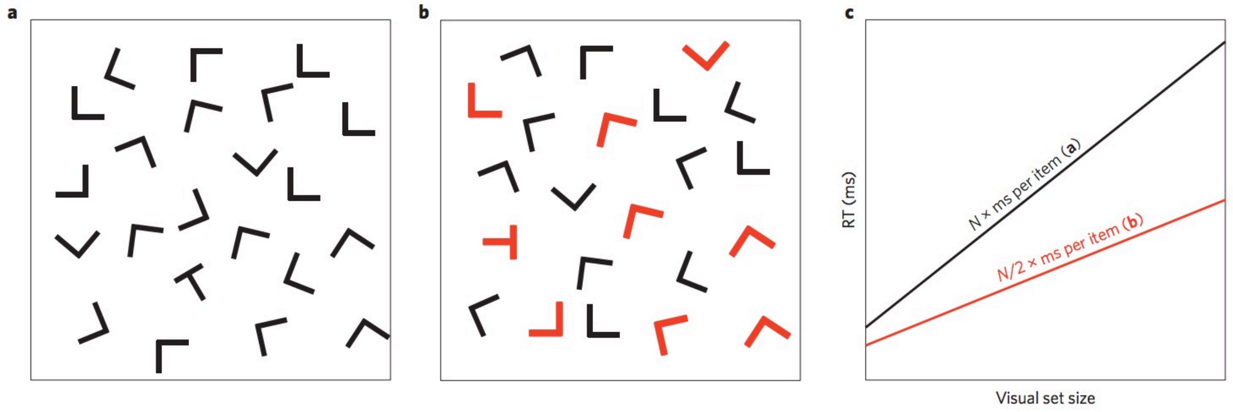


Figure 2. A simple visual search task paradigm. **a)** Feature Search; the letter T (the target) is presented among a series of Ls (distractors). **b)** Conjunction search; the same task as (a) but with a red T target and half as many distractors sharing the color feature of the target; black letters can be ignored without determinant to the efficiency of the search due to color being a selected search feature. **c)** The efficiency of the search is represented by the slope—relating reaction time (RT in milliseconds) to the set size in the

visual search space—of the functions for visual search tasks (a) and (b). Reprinted from "Five Factors That Guide Attention in Visual Search" by Wolfe and Horowitz, 2017, *Nature Human Behavior*, 1(3), 1-8.

Visual search becomes interesting when used as a tool for determining how attention is applied to a search space. A red T among black Ts is easy to find (referred to as "feature search"; like Figure 2a, which uses shape rather than color as the distinguishing feature) because the relevant feature pops out. Figure 2b is referred to as conjunction search because it requires the integration of more than one feature (here, color and shape) in order to distinguish the target, and thus much more time. The feature integration theory (Treisman & Gelade, 1980) attempts to explain these results by proposing that certain visual features (e.g., color, orientation, luminance, motion, speed) are registered in a rapid, automatic way using preattentive processes; but when integrating multiple features—like in conjunction search—a later, attentional process is recruited which integrates these features and binds them together into a singular object by which they can be coded serially via focal attention. Further evidence for this theory has been put forth based on conjunction illusions, in which features are incorrectly bound to objects. For example, if a green O and a red X are flashed on a screen so quickly that these later processes (serial search/focal attention) do not engage, an observer is likely to report having seen a red O and a green X (Treisman & Gelade, 1980). Similar illusions occur outside of the visual domain as well (e.g., in the auditory domain; Thompson, Hall, & Pressing, 2001), suggesting that this process is "feature-independent." It becomes clear why visual searches are such a valuable tool: By varying factors of the search space—e.g., frequency of features, selection history, etc.—or adding factors into the search space—e.g., spatial cues, rewards—one can determine much about the human attentional system and distinguish the processes at play in attention. But what do these distinctions really mean? What can they tell us?

One thing that these distinctions inform are the uncovering of the mechanisms of these processes: Selection is chiefly achieved via a system called the frontoparietal attention network, which actively adjusts the susceptibility of neurons in perceptual areas of the brain (e.g., the visual cortex; Silvanto, Muggleton, Lavie, & Walsh, 2009; Taylor, Nobre, & Rushworth, 2007). This constructs a kind of priority map on "top" of extracted perceptual features (such as color features or spatial frequency in the visual domain; see Figure 3). It's important to note that this frontoparietal "attention" network "map" is not specific to vision or sensation, but is instead thought to be multisensory (feature-independent) and to integrate both top-down and bottom-up inputs—anatomically distinct sub-networks (see Figure 4; Ptak, 2012). Paradigms such as the visual search task allow the characterization of brain dynamics and reveal how the brain represents information and maintains attentional processes.

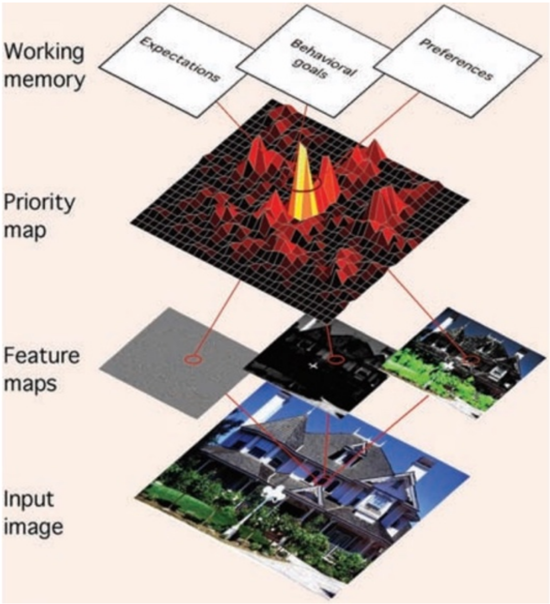


**Figure 3.** A simple model of the computation of attention. A sensory-domain area—here the visual cortex—automatically (pre-attentively) deconstructs an input into a series of feature maps, representing different aspects of the input (we see in the above picture spatial frequency, intensity, and color—deconstructed aspects of the image). The priority map then integrates these feature maps and combines them with top-down information (held in working memory), which reflects top-down orientations such as goals, plans, and expectations. Selection of spatial priority therefore represents the interaction of functionally distinct top-down and bottom-up processes. Reprinted from "The Frontoparietal Attention Network of the Human Brain: Action, Saliency, and a Priority Map of the Environment" by Ptak, 2012, *The Neuroscientist, 18*(5), 502-515.

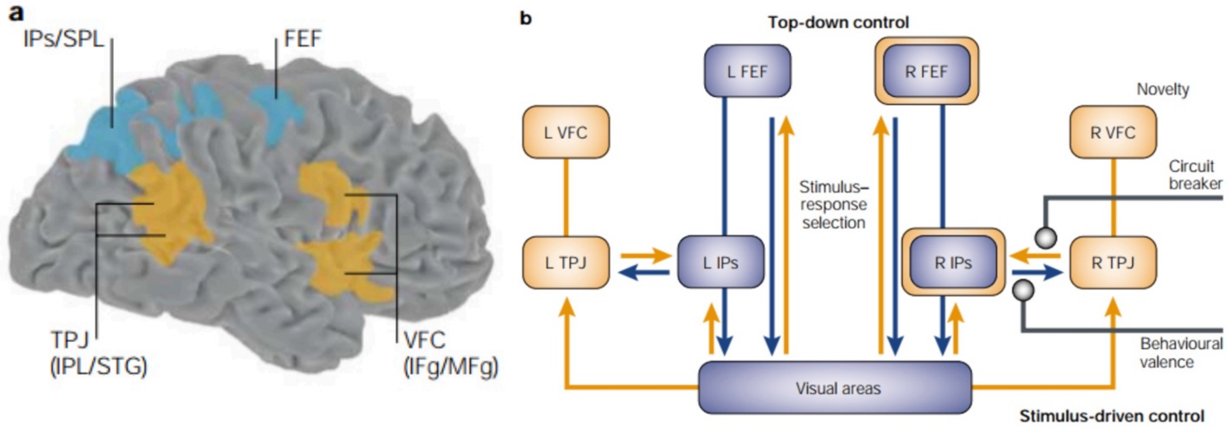


**Figure 4.** An anatomical model of attentional control. a) Dorsal (blue) and ventral (orange)

frontoparietal network regions in the human brain: The dorsal network is thought to mediate top-down processes such as working memory, intentions and more abstract reasoning, while the ventral network represented here indicates stimulus-driven, bottom-down processes. b) Model of the functional interaction of these two anatomical systems. IPs and FEF are modulated by stimulus-driven control originating from the ventral network; TPJs and IPs can block on-going top-down control processes when salient and unattended stimuli are detected in some sensory domain. IPL, inferior parietal lobule; IPs, intraparietal sulcus; FEF, frontal eye field; IFg, inferior frontal gyrus; L, left; MFg, middle frontal gyrus; R, right; SPL, superior parietal lobule; STG, superior temporal gyrus; TPJ, temporoparietal junction; VFC, ventral frontal cortex. Adapted from “Control of Goal-Directed and Stimulus-Driven Attention in the Brain” by Corbetta and Shulman, 2002, *Nature Reviews Neuroscience*, *3*(3), 201-215.

While few deny that uncovering the neural bases of processes such as attention is interesting, many might claim that doing so has few applications, or reveals little other than associations between brain states and experience (such as the experience of performing a visual search task). But the cognitive sciences have much to unravel, and subsequently, much to apply, from learning about localization in the brain. This has certainly proven true in research of the other primary component of learning yet to be discussed: Memory.

### ***1.3 Memory***

Despite the omnipresence of memory in daily life—remembering how to fill the coffee pot in the morning; recalling distinct sensory details of events that happened decades ago; or even predicting novel, future events (Shacter et al., 2012) —the rigorous study of memory is a relatively recent development. As such an obvious and essential function of the brain, why is it that memory so eludes inspection? Part of the problem is that the brain is always adjusting itself to new information; just by reading this sentence, your brain has been changed ever-so-slightly. Thus memory is a vastly complex, ephemeral system. How does one study something whose process is so hidden—physically and conceptually (ineffable?)—and whose state is in constant flux?

The word “memory” also has many different functions and meanings. Memory might refer to (1) the location where information is stored, (2) the object that contains the content of experience (like an *engram*; more on this soon), or (3), a process by which one acquires, stores, or retrieves information to be utilized such that it may be available later to inject into the current mental state (Radvansky, 2015). Taking a dive into the historical roots of attention research might provide some clarity on how memory is defined in modern day psychology and neuroscience.

To Plato, memory acted as the bridge between the perceptual world and the world of the abstract (Radvansky, 2015). He likened it to a block of wax in the soul upon which perceptions and thoughts may be imprinted, but also by which memory might be “rubbed out” of existence (Bernecker, 2010). But it wasn’t until Quintilian, in his treatise *On the Education of the Orator*, that the history of memory research alluded to the idea that memories take some non-zero amount of time to form proper—a “process of ripening and maturing” that can happen overnight. This is the only known reference to the process of memory consolidation—wherein memory traces become stabilized—until the late 19th century (Polster, Nadel, & Schacter, 1991, p. 96). It seems that the earliest forms of studying memory relied heavily on metaphor—and it might make sense that the preserved history of the study of memory is delivered in this way because metaphors can make statements more likely to be remembered as well as more likely to be considered reliable, and therefore disseminated; Read, Cesa, Jones, & Collins, 1990). It was out of this idea of memory consolidation that much of what we know about memory in the brain began to unfold.

Ribot, among those studying memory at the turn of the 19th century, invoked the idea of memory consolidation to explain the retrograde amnesia of a brain trauma patient of his by stating that “in order that a recollection may organize and fix itself, a certain time is necessary, which [in this case of trauma] does not suffice” (Ribot, 1892, p. 799). It was after this, in the early 20th century, that a productive and fierce debate oriented around forgetting began; it was revealed, non-intuitively perhaps, that forgetting is in a large part due to memory interference rather than simple decay (e.g., McGeoch, 1932)—though more modern perspectives on this issue propose contributions and functional interactions of both systems during and after consolidation processes (e.g., Altmann & Gray, 2002). As behaviorism become a strong force in the field of psychology and internal investigation waned, many basic questions about consolidation remained unprobed until the early 40s, during which Zubin and Berrera (1941) presented a paradigm for systematic investigation into consolidation via electroconvulsive shock (Polster et al., 1991). Their discovery that by shocking the brain, one could disrupt consolidation processes had a large impact on the psychological world. This renewed interest in memory processes gave rise to a more internally-focused movement in the era of behaviorism.

The work of Lashley in the mid-20th century also assisted this post-behaviorist attitude by bringing the question “where are memories located?” back to the psychological forefront. By parametrically lesioning rats’ brains and subsequently running them through mazes designed to test their spatial memory, Lashley ostensibly appeared to have ended the debate: No matter *where* Lashley lesioned rats’ brains, they were mostly still able to remember how to move through their maze; it was, rather, the proportion of brain matter removed that best predicted rats’ memory failures (see Figure 6). This lead to the evolution of two neuroscientific hypotheses: the mass action hypothesis—the notion that the whole brain participates in every behavior; and the principle of equipotentiality—which refers to the apparent capacity for functions of damaged regions to be adopted by undamaged regions; and additionally to the conclusion that the engram—matter in which a memory is stored—lacks a specific location. [*Where* a memory is located was thus perhaps an ill-posed question—“perhaps” because “where” might not be a singular location (or perhaps it can be; see McCormick et al., 1981)]. This seemed entirely contrary to the conclusions of neurological researchers at the end of the 20th century, who claimed that functions were segregated in space within the brain (e.g., Broca and Wernicke, see section 2.3). How can these two seemingly incompatible accounts of how the brain is organized be reconciled?

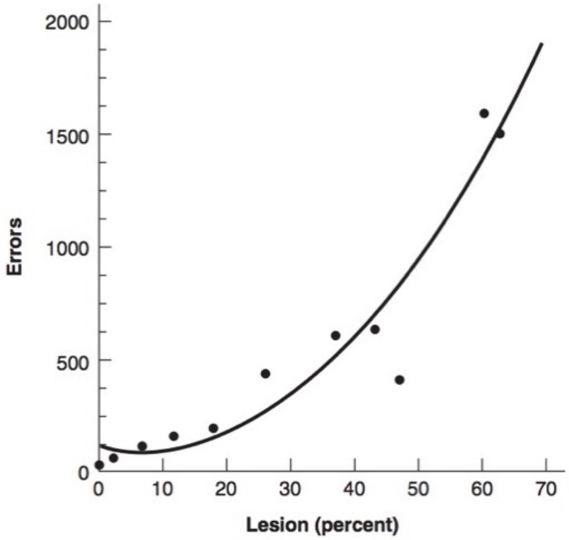


**Figure 6.** The results of Lashley's search for the engram; memory errors increase exponentially relative to the percent of brain lesioned. Reprinted from "Human Memory" (2nd edition) by Radvansky, 2015, *Routledge*.

On the shoulders of Lashley, psychologists at the midpoint of the 20th century were committed to the account of memory as a distributed system in the brain. At the time, the cognitive faculties were divided functionally by types of intellect and perception: There was a verbal system, a visual system, an auditory system, etc.; and the memories for each of these processes (e.g., "verbal memory") was stored in these distinct brain areas (Eichenbaum, 2013). But all of that was about to change with patient H.M.

At the same time that the cognitive revolution in psychology was beginning to simmer, Scoville and Milner reported, in 1957, the profound mnemonic side-effect of a rare and experimental surgery performed to abate severe epilepsy—a bilateral medial temporal lobe (MTL) resection—of patient H.M. (1926-2008; see Figure 7). Little did Milner know that this case study would become a foundational investigation that would rocket the neurosciences through the cognitive revolution and beyond.

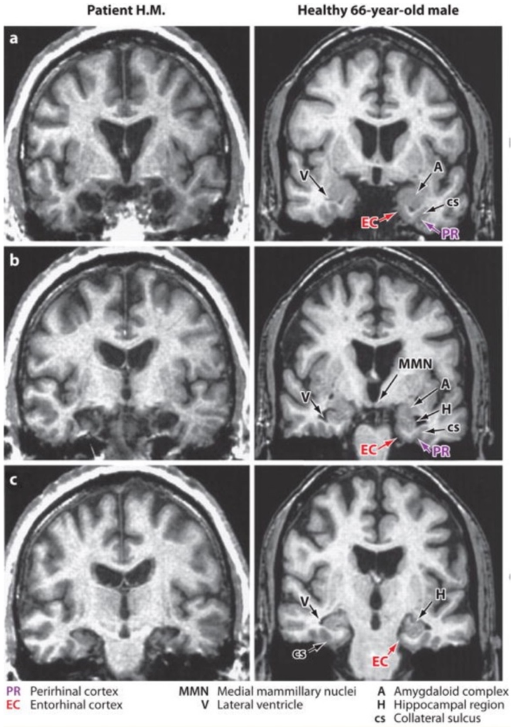


**Figure 7.** Magnetic resonance images compare the temporal lobes of patient H.M. (at age 67) to those of a comparable subject (66-year-old male). Structures crucial for proper memory encoding were removed from H.M.'s brain in an experimental surgery; they are highlighted by the labels listed above. The images move from rostral (a) to caudal (c). Reprinted from "The Cognitive Neuroscience of Human Memory Since H.M." by Squire and Wixted, 2011, *Annual Review of Neuroscience*, 34, 259–288.

H.M., if you were to have met him, would have appeared fairly normal. In fact, you might even have been able to have a conversation with him without noticing his impairment, as long as you didn't reference the conversation itself. That is, after the surgery, H.M. developed profound anterograde amnesia—he forgot daily events nearly as fast as they occurred—but maintained his general intelligence, perceptions, etc. (Squire, 2009). Additionally, the memory impairments were global, meaning that the deficits were evident across all modalities of cognition (e.g., his memory for auditory things was just as impaired as his memory for visual things; Eichenbaum, 2013). These findings (reported in Scoville & Milner, 1957) shattered the world of memory research: Memory was promoted from an integrated feature of modalities to a distinct cerebral function, and, at least in part, a distinct locale; specifically, this finding introduced the idea that the brain has separated its perceptual and cognitive areas and functions from the system that maintains memories for the engagement of these systems (Squire & Wixted, 2011).

But this wasn't all that H.M.'s situation told us about the brain. First, while H.M. wasn't able to acquire new information very well, Corkin (1968) showed that H.M. had no trouble acquiring novel motor skills; so motor skill acquisition occurs—at least predominantly—elsewhere than the medial temporal lobe. Second, H.M. was able to sustain his attention well and maintain memories for items for short periods of time; so the resected areas were not overwhelmingly responsible for immediate memory or for rehearsal and maintenance of what is now known as working memory. Third, H.M. had good access to facts and (at least appeared to have "good" access to; see Moscovitch et al., 2005) events from the time before his surgery; so, the medial temporal lobe can't be the sole or chief area for the storage of long-term memories either—that function has mostly been attributed to neocortical regions (Squire & Wixted, 2011) as well as to a diversity of supporting networks (which support and maintain, e.g., classical conditioning and nonassociative learning; see Figure 8). What H.M. could *not* do also uncovered a previously unsegregated type of memory called declarative memory, referring to the acquisition of explicit facts and events (see Figure 8; Squire, 2009; Squire & Wixted, 2011; Eichenbaum, 2013).

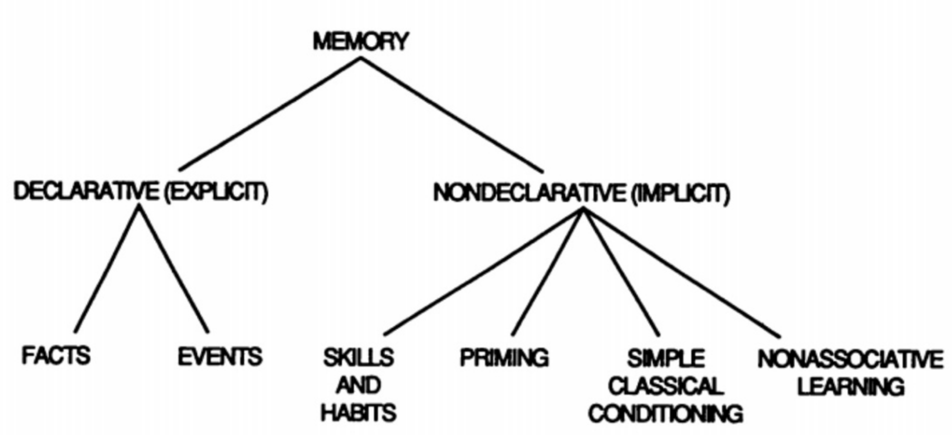


**Figure 8.** Taxonomy of memory. Declarative memory refers to the conscious remembrance of facts and events; nondeclarative memory references the group of memory types that alter behavior unconsciously and are independent of the medial temporal lobe—they include skills and habits, priming, classical conditions, and nonassociative learning. Reprinted from "The Medial Temporal Lobe Memory System" by Squire and Zola-Morgan, 1991, *Science*, *253*(5026), 1380–1386.

Each of these ideas has generated a substantial body of literature and informed the progression of the relevant sub-fields of psychology and neuroscience as well as both as a whole. For example, later research (Squire & Zola-Morgan, 1991), spurned by this discovery, identified many functional (see Figure 9) components of the medial temporal lobe memory system—consisting of the hippocampus as well as the bordering perirhinal, entorhinal, and parahippocampal cortices (see Figure 10). But while it would seem then that each of these systems should now be clearly delineated, active research since H.M. on this basis has postulated that the medial temporal lobe can be involved in other functions beyond declarative memory, such as visual perception, working memory, and forms of spatial cognition; though the most robust effects are still in the declarative memory realm (Squire & Wixted, 2011).

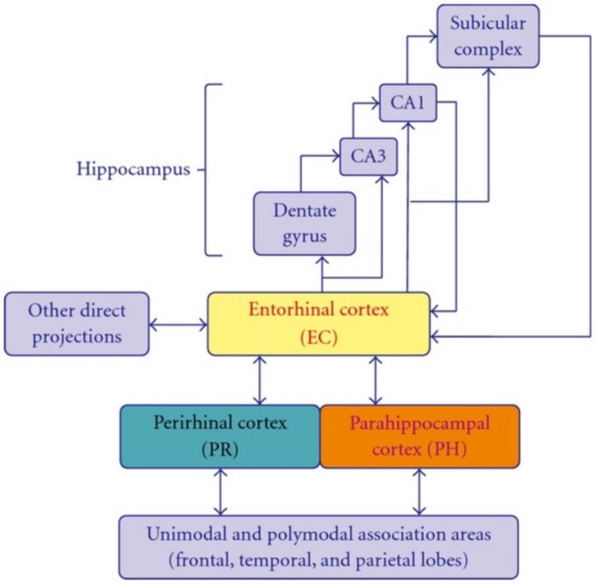


**Figure 9.** A diagram representing the connections of the medial temporal lobe memory system for declarative memory. Reprinted from “Role of Functional MRI in Presurgical Evaluation of Memory Function in Temporal Lobe Epilepsy” by Limotai and Mirsattari, 2012, *Epilepsy Research and Treatment*, 2012, 1-12

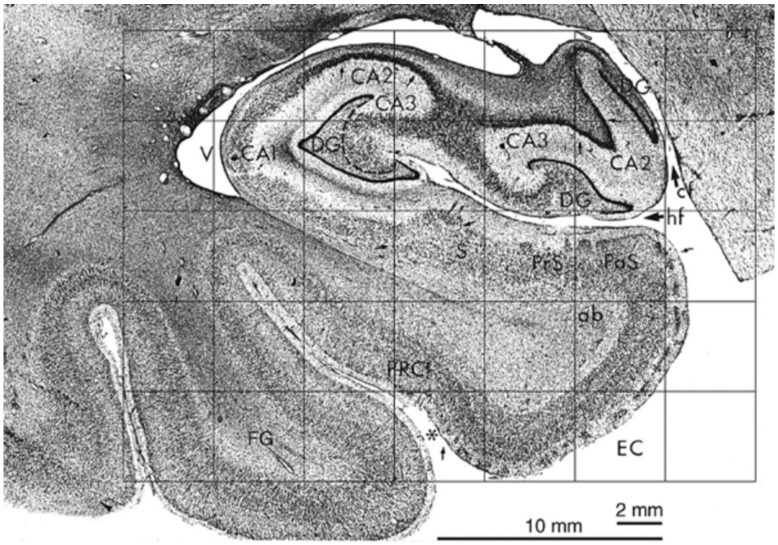


**Figure 10.** Stained (Nissl method) coronal section of the medial temporal lobe (at the

caudal aspect of the uncus). ab, angular bundle; CA2, CA2 field of the hippocampus; CA3, CA3 field of the hippocampus; cf, choroidal fissure; DG, dentate gyrus; EC, entorhinal cortex; FG, fusiform gyrus; hf, hippocampal fissure; PaS, parasubiculum; PRC, perirhinal cortex (areas 35 and 36); PrS, presubiculum; V, temporal horn of the lateral ventricle. Reprinted from “The Hippocampus Book” by Andersen et al., 2007, New York: Oxford University Press.

The legacy of H.M. continues to define contemporary debates about the bases of memory. While the results of H.M.’s (and rare few other patients’) bilateral MTL lesions informed researchers as to the role of the medial temporal lobe—particularly the hippocampus—in many operations, such as *recall* (the capacity to retrieve memories sans any cue), disputes rage on about potentially more nuanced and complicated processes in the brain: For example, as to whether *recognition* of past information is impaired in such a circumstance of lesions (Lacot et al., 2017).

Recognition involves assessing whether or not, or to what extent, a stimulus has been previously experienced. A contentious topic in contemporary memory research revolves around the question of whether or not this cognitive function is underpinned by a single- or dual-process system, and where they are localized. The dual-process account suggests that recognition is the result of two, at-least-partially independent systems: Familiarity, referring to—put colloquially—the distinct *feeling* of having experienced the object before; and also recollection, typically conceived of as a slower, more effortful occurrence that brings to conscious mind information about the prior occurrence of the object of recognition and the contexts in which it has previously appeared. According to those in favor of a single-process account, assessments as to whether or not the subject has experienced the current information are attributed solely to the familiarity system (Squire & Wixted, 2011; Gardiner & Parkin, 1990; Rugg & Curran, 2007). While still a hot topic, most of the research has fallen on the side of the dual-process accounts; and on the notion that these mnemonic operations are hippocampally mediated (Rugg & Curran, 2007; Squire & Wixted, 2011, but cf. Slotnick, 2014 or Slotnick & Dodson, 2005).

### *1.4 Interactions of Attention and Memory*

While it's easy to separate attention and memory into sections of a paper, it's much harder to separate them in the brain. Even when one's mnemonic and attentional goals are in conflict, there seems to be something inseparable about the two systems (Chun & Turk-Browne, 2007). Why so?

For example, in attempting some mnemonically-mediated processes, attention may be a hindrance: Moments of insight ("Aha!") in which it would appear that the answer to a problem suddenly arises to consciousness without being attended ("Eureka!"), may be facilitated by a *lack* of attention. Evidence for this phenomenon comes from research showing that—relative to healthy controls—those who lack attention-mediating brain areas are actually much better at solving some types of difficult (i.e., non-intuitive) problems which require relaxations of particular constraints imposed upon the parameters of the problem based on memory for the function of those items (Reverberi, Toraldo, D'Agostini, & Skrap, 2005). The relevant researchers refer to this phenomenon as the result of a deficiency of attentional processes which "sculpt the response space"—processes which identify the context-reasonable responses to problems from those that do not fit the problem. For example, we don't build our houses of glass because we *know*—we have a memory for the fact that—glass breaks easily.

Further evidence (Metzler, 2001) for this relationship has shown that left frontal cortex damage can eliminate semantic priming effects. [Semantic priming refers to the automatic context evaluation of responses; for example, the word "car" is identified more quickly or accurately after the presentation a related word—like "wheel"—relative to an unrelated word—like "fork"]. Likewise, Thompson-Schill and colleagues (1998) demonstrated that similar damage can reduce one's abilities to generate verbs that are appropriate for nouns, suggesting that the "wrong" (i.e., context-nonspecific) information—as determined by a mnemonic categorization process—is more likely to be selected when one has diminished attentional control.

Additionally, for some attention-mediated processes, memory can be a hindrance to one's goals; such an interaction may be evidenced by the Think No-Think paradigm (Anderson & Green, 2001), which requires participants to either think or not think about an associate of an item. For example, having studied "wolf" and "chicken" as a cue-and-response pair, when presented with "wolf," one either tries to think of "chicken" (think) or tries their best to not think about "chicken" (no-think). This work has suggested that attention may not solely be applied to the external environment, but also the internal one, demonstrating that attentional processes can facilitate or inhibit mnemonic ones: In the process of trying to *think* about something, one's attention is oriented towards their memories and top-down attentional processes (primarily via the dorsolateral prefrontal cortex; Anderson et al., 2004) might increase activity in the hippocampus such that the response would be “picked out” more easily; while in the process of *not* thinking about something, one’s attention could be oriented away from their memory (and, say, towards the perceptual features of the word cue on the screen) and this top-down activity may actively *inhibit* activity in the hippocampus, drawing ones attention away from the response when cued. This inhibitory exemplar is thought to induce a "suppression mode” of the hippocampus, in contrast to an effortful direction of attention towards retrieval—which would produce a “retrieval mode” (Hulbert, Hirschstein, Brontë, & Broughton, 2018; Rugg & Wilding, 2000).

The AtoM (Attention to Memory) hypothesis, put forward by Ciaramelli, Grady, and Moscovitch (2008) attempts to align with these and other dynamical interactions of attention and memory. AtoM is a dual-process theory which holds that the top-down system mediates strategic retrieval—which directs mnemonic searches, constrains the search space, and reinstates material that is goal-relevant in response to irrelevant events so as to monitor and verify the products of retrieval. The second, bottom-up, “attention-to-memory” process mediates direct retrieval, capturing attention when memory contents retrieved by the medial temporal lobe match a current event (as a target might capture one’s attention in visual search space; see Cabeza et al., 2011). Examples of the former might occur if an individual is not confident in a memory or when current events are similar, but not identical to a memory (i.e., when discrimination is difficult or stimuli ambiguous); examples of the latter might occur when one is confident and when memories are rich, strong, and/or match current events well.

Other interactions of attention and memory are demonstrated when examining the relationship between attentional burden and memory processes. Divided attention has been shown to disrupt effortful and elaborative encoding, but have little effect during retrieval of that information (Craik, Govoni, Naveh-Benjamin, & Anderson, 1996). It's thought that additional attentional costs in phases of encoding may reflect the entering of an encoding mode—a state associated with preparing and maintaining intentional encoding processes—which requires sustained attentional resources. This is perhaps due to the need to maintain encoding procedures and transfer them into working memory (Naveh-Benjamin et al., 2007). However, recent findings have demonstrated that divided attention while encoding has another consequence related to the reduction of attentional control capacity: increased encoding of non-target items (which would be expected to create interference between the target items and any encoded distracters; Weeks & Hasher, 2017). Furthermore, these effects may be present under both intentional and incidental learning instruction conditions.

Intentional encoding refers to when one intends to remember an event for later, while incidental encoding happens often, referring to a more passive acquisition of memory for events. While, intuitively, intentional encoding can facilitate better memory performance (e.g., Neill et al., 1990), this is not always the case. For example, visual information tends to be equally able to be recognized in incidental and intentional encoding paradigms (Castelhano & Henderson, 2005; Bird, 1976). Furthermore, implicit memory for unattended, novel shapes can be preserved over long delays, yet explicit memory for these objects might be entirely unavailable (DeSchepper & Treisman, 1996).

Within encoding, incidental and intentional processes produce different brain activity patterns as well (Bernstein, Beig, Siegenthaler & Grady, 2002); however, previous studies have often found no effect between incidental and intentional learning on brain activation in recognition phases (van der Veen et al., 2006). While differences in activations *can* predict successful recognition in some paradigms, overall it appears that differences in encoding strategies often have stronger observable effects on brain activity during encoding than during recognition (Bernstein et al., 2002).

Furthermore, correlates of successful vs. unsuccessful incidental memory formation have also suggested distinct brain activation patterns and a non-intuitive involvement of attention in incidental encoding. Clemens and colleagues (2015) demonstrated that those who exhibited greater incidental memory performance showed greater activity in attentional (and salience-detection) brain areas. Thus even in incidental encoding, attentional processes may play a crucial role in remembering.

As we've seen, answers to questions about localization and time course of function assist in determining how the brain operates on and represents information (and its implications are much debated; e.g., Bennett, Dennett, Hacker, & Searle, 2009). This paper has discussed how lesions and patient studies (and even electroconvulsive shock) can help answer this question; but much of the research referenced from recent years has become increasingly reliant on and interested in using advanced hardware and computational techniques to elucidate the role of cognition in the brain. The following section explores the technical methods employed in these contemporary psychological/neuroscientific investigations and their physiological bases.

## Section 2

### *2.1 Psychological Technology*

Psychology was once limited to studying mental processes through self-examination or by observing behavior through one's eyes. A survey of these historical foundations of psychology might summon other intellectuals from the past, like Descartes, who employed introspection to uncover mechanisms of the mind; or Ebbinghaus (1885/1913), whose famous forgetting curve was derived solely from his own behavioral self-experimentation.

Before this time—the formalization of psychology—studying the mind was attributed to paranormalism and the occult, feared for its agents' "mind-reading" abilities and their purported connections to evil spirits or demons (Benjamin, 2000). Now, uses of contemporary technology in psychological inquiry have largely supplanted this image: Colorful models of brains-in-function, high-resolution images of in-vivo neurons, and a resulting set of applications unimaginable but fifty years ago typify its popular depictions today. The arcane and suspiciously-viewed psychology of the past now touts scientific rigor; its cultural "connections" to paranormalism in media and culture have been largely supplanted by ones of materialism.

Now, anchored by the powers of new technology, the capability for actual mind-reading may be closer at hand than ever before. Contemporary science-fiction writers have sculpted great and terrible worlds that result from the clever or misuse of such capabilities. But how far, exactly, have the advancements of this technology truly taken us in mind-reading? To where do they promise?

### *2.2 What Is Mind Reading?*

Sarah and Greg go to the same school. One day in the library, as she sits next to him, Sarah recognizes that Greg's face appears more blushed. If Sara were to infer that Greg has a crush on her—purely based on this physiological information—is that mind-reading? Now, if a computer had access to every face in the world, at all points in time, and were to analyze the color spectrum of Greg's face based on these "training data" and come to the same conclusion as Sarah with the same degree of certainty, would the answer to that question be the same? In fact, Greg is just anxiously trying to learn a whole semester of content before his test begins in seven hours—did Sarah fail to read his mind? Would the computer have failed to read his mind as well?

"What is mind-reading?" seems like a remarkably simple question at the outset. But in attempting to draw a clear line, the definitions blur. How does one even begin to measure a mind's actions?

### *2.3 Functional Localization*

Today, one might take for granted the idea of functional localization—the notion that different brain regions have distinct functions. While it was not always known that the brain is segregated by function, even in ancient Greece (where the soul was variously thought to reside in body parts such as the heart and the liver) intellect had been associated with the head (Zola-Morgan, 1995). Indeed, the organization of the nervous system was first uncovered by Greek anatomist Galen (130-200 AD), who, using animal models, determined many things about how the brain operates: that the brain was the source of sensation and thought; the controller of movement; and that the spinal cord was an extension of the brain, conducting sensory signals from nerves and sending motor signals to muscles. Rather than expanding on these findings, however, later generations in Europe "accepted as undisputed and indisputable his views in every branch of medicine" (Gross, 1987, p. 844) and Galen's ideas dominated thought around the functioning of the brain for 1500 years.

While some researchers (e.g., Thomas Willis and Robert Whytt) pushed the boundaries of neurological knowledge in the 17th-18th centuries, most of the debate in Europe was influenced by the leanings of a series of powerful religious groups and centered around whether mental functions derived from brain matter itself or the space in the brain's ventricles (which gained significant following because empty space was more conceptually infinite, and God's role in the brain was thought to be too limited by material existence). It was not until the 19th century that those who studied the brain were pursuing useful questions about localization (Gross, 1987).

Franz Joseph Gall, whom many claim as the founder of the school of the pseudoscience "phrenology" (Gall is often associated with the phrenology movement—e.g., Tyler, 2014; Berker, Berker, & Smith, 1988; but *cf.* Zola-Morgan, 1995), was among those most prominent who put forward evidence and advocated for functional localization in the brain. Gall attempted to define the many functions of the brain according to specific morphological features of the skull and the shape of the head (Eichenbaum, 2011). While this might have ultimately been a fairly fruitless endeavor (spoiler: There's no real relation), Gall and others did enough to get the ball rolling in the right direction.

By the 1870s, studies utilizing brain stimulation on dogs (that had, crucially, much larger cortices than other animal models previously used to investigate functional localization—like birds; Eichenbaum, 2011) began to investigate the localization of motor areas. Fritsch and Hitzig were the pioneers of this model. They showed that stimulating zones in the frontal cortex resulted in specific muscle movements and that by adjusting the locus of stimulation, they could determine a relative "map" on the cortex corresponding to different motor areas (e.g., "zapping" an anterior-lobe area might produce movement of the forepaw on the contralateral side, while doing so to the posterior-lobe area might produce the same movement on the adjacent side; Eichenbaum, 2011). While this evidenced the idea of functional localization, and further substantiation was put forth by Ferrior (using monkeys to ascertain the loci of sensation processing; Heffner, 1987), the questions as to whether "higher order" functions were localized in humans had yet to be confirmed.

Around the same time, Broca, a French physician, reported a patient with a severe condition that prevented him from speaking words, yet the patient's mouth was not affected and he was able to understand speech perfectly well. The patient was known as

"Tan" for the sound of the vocalizations he produced (real name: Labourgne). Upon Tan's death (unrelated to the neurological condition in question), Broca performed an autopsy and discovered a specific area of neurological damage on the third convolution of the left frontal lobe (Broca, 1861/2003). The case report provided additional support for the idea of localization in the brain, and its association with the selective behavioral disorder was compelling evidence for the functional localization of "higher functions." Further substantiation was provided when Carl Wernicke discovered a complementary case in 1894. This patient was not impaired in any relevant modality other than being severely impaired in the ability to comprehend speech; in line with what Broca had uncovered, Wernicke found neurological damage restricted to a nearby, but morphologically distinct brain region in the left temporal cortex (Eichenbaum, 2011) (see Figure 11).

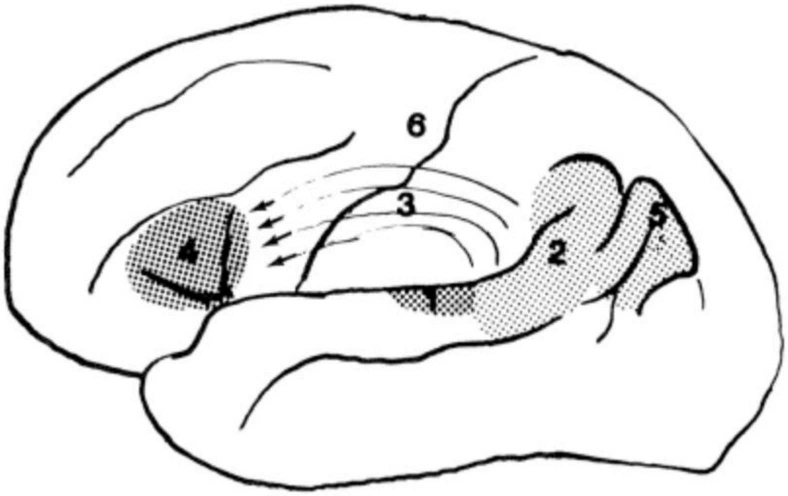


**Figure 11.** Diagram of the left cortical regions functionally associated with verbal language. Broca's area (4) and Wernicke's area (2) are still the terms used to refer to the regions responsible for the majority of speech production and comprehension, respectively. 1, Heschel's gyrus; 3, arcuate fasciculus; 5, angular gyrus; 6, motor cortex. Reprinted from "Disorders of the Nervous System" by Reeves, 2008, dartmouth.edu.

Historically, neuroscience has derived much of its foundation from the study of those with damage to particular brain areas. But these methods of the past are limited in both their accessibility and scope: There are few people who lack specific anatomical regions of the brain, and frequently those who are missing particular regions have further damage elsewhere (often due to the cause of their relevant neurological condition). Even conclusions drawn from research on H.M. have suffered from the potential of a concealed prefrontal lesion that had previously gone undetected, throwing much of the data collected with H.M. into question (Annese et al., 2014). Furthermore, regulation resultant from ethically dubious research, a greater understanding of the import of the brain as a whole, and advancements in treatment modalities have discouraged surgical methods of "correcting" neurological disorders or disordered behavior. So while this technique has proved fruitful in the immediate sense of providing a platform of functional localization, it’s quite clear that it would be advantageous for new methods to evolve for studying brain function.

Much of what is known in modern times about localization in the brain has arisen quite recently due to rapid development of psychological technology, most prominently functional magnetic resonance imaging (fMRI). Today, functional magnetic resonance imaging (fMRI) is used to measure brain activation patterns and is perhaps the most well-known modern psychological technology; it was only first introduced in the 1990s and yet accounts for at least 40,000 published papers (Eklund, Nichols, & Knutsson, 2016). Is fMRI simply “the new phrenology” (Uttal, 2002), associating psychological states with brain states? Or does fMRI hold the secret to mind-reading?

### *2.4 fMRI*

Magnetic resonance imaging (MRI; sans “functional”), only introduced 20 years prior to *f*MRI, is a technology primarily used to produce images of the internal body (organs, bones, muscles, brain, etc.). By inducing a strong magnetic field, scientists (chiefly Isidor Rabi) in the hay-day-of-quantum 1930s discovered that atomic nuclei absorb and emit electromagnetic radiation, and that they do so at a resonance specific to the type of atom—the “Larmor Frequency” (Buxton, 2013).

The nuclei of atoms—the ones that are capable of being imaged with MRI—normally spin around an axis due to their magnetic properties, kind of like the earth. But the orientation of this spin is different across atoms in a sample (e.g., the body) because they aren’t aligned in nice little rows and columns, they’re floating in 3 dimensional space. When a magnetic field is applied to an object, however, some of the nuclei of the atoms in that object (i.e., those that are magnetically active; those that possess spin; those that are visible with MRI) are affected such that the movement of their nuclei align with the magnetic field (along a consistent vector) (Buxton, 2013; Goense, Bohraus, & Logothetis, 2016). Additional energy (i.e., radio frequencies) can then be directed at these magnetized nuclei such that it is absorbed by these aligned nuclei. The nuclei are thus slightly deflected from their magnetic vector because of this energy, so when the radio emission is turned off, the nuclei return to a resting state and re-emit electromagnetic frequencies specific to the magnetic properties of the isotope of the atoms (i.e., specific to the altered spin). This produces a measure of the kind of atom present at that point in space. Conventional MRI uses hydrogen atoms due to the high presence of this element in water and the even higher presence of water in the human body (Sokoloff, 2008). By applying a gradient magnetic field and using computers to rapidly complete this process through space, this technology became able to create the meaningful images of the internal body one sees at the doctor's office today.

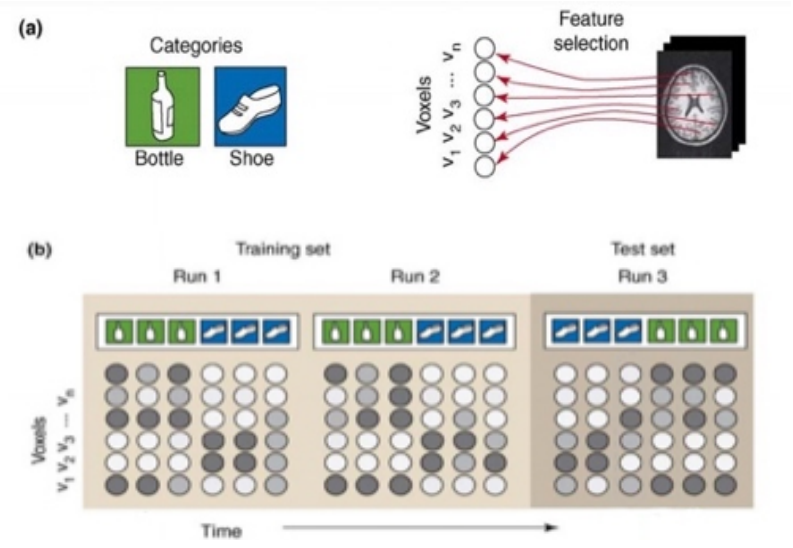
One can imagine the impact such a technology might have had, especially on the medical community, given its ability to image in-vivo and its distinct advantages over x-ray and CT scans that involve ionizing radiation (Buxton, 2013). One can also imagine that to psychologists (or philosophers, etc.) hopeful that imaging the in-vivo brain using MRI would uncover secrets of the mind, the technology's limits as to such might have been disappointing: MRI was revolutionary for neurosurgeons, for example—it allowed faster identification and diagnosis of tumors, abscesses, aneurysms, hematomas, degenerative diseases, etc. based on morphology (Silva, See, Essayed, Golby, & Tie, 2017)—but ultimately, these scans alone tell very little about how the brain functions. Thus (and in complete acceptance of retrospective bias) an *f*MRI—avec "functional"—seems the logical next leap

Crucially, fMRI capitalizes on the physiology of neurons to image their "activity" throughout the brain as well. The crucial step in developing a functional MRI came with the discovery that the magnetic properties of hemoglobin—the molecule in blood that carries oxygen throughout blood—change when it becomes deoxygenated into deoxyhemoglobin as it does when neurons use the oxygen (Sokoloff, 2008). Because blood is recruited by neurons in their activation, fMRI can measure correlates of activity of groups of neurons by imaging blood flow changes, or more precisely, measure oxyhemoglobin to deoxyhemoglobin ratio increases associated with neural blood recruitment. These changes in recruitment of oxygenated blood that fMRI detects are referred to as blood-oxygen level dependent (BOLD) signal (Ogawa, Lee, Key, & Tank, 1990; Goense et al., 2016). Experiments performed by Sieji Ogawa and colleagues in the early 1990s were a critical leap for functional brain mapping, demonstrating that noninvasive neuroimaging can distinguish BOLD activity (Ogawa et al., 1990). This technology has become a staple of neuroscience because the data it provides are relatively spatially precise, enabling researchers to image where processes tend to occur in the brain. Typical brain-voxel resolution in fMRI is 3x3x3mm, which might sound small, but this resolution still sums over large populations of neurons. Even with recent neuroimaging advances—contemporary "high-resolution" fMRI is considered sub-millimeter (Goense et al., 2016)—one cubic millimeter may still contain over one million individual cells (deCharms, 2008). But despite imperfections, fMRI offers a fairly rapid and non-invasive approach to functional brain assessment and continues to pave the way alongside contemporary psychological inquiry.

While fMRI is useful in delineating functional localization in the brain, it might appear to an outside viewer that the applications of the technology end there, but this is far from true. For example, while preliminary, fMRI techniques are of increasing interest in the domain of diagnostics. Patients with disordered cognition, mood, etc. may one day be able to be identify or even diagnosed without need for potentially slower behavioral assessments, or fMRI might reduce ambiguity or disagreement via traditional clinical diagnostics (Melo et al., 2017); this has become a cutting edge field with the development and application of deep learning methods to process neuroimaging data (Wen et al., 2018), especially in identifying Alzheimer's disease. fMRI used in combination with other investigative tools, such as genetics, pushes the boundaries of its capabilities (Matthews, 2006). Furthermore, presymptomatic diagnostics offers the potential to recognize disordered brain function before problems even present themselves in patients' daily life, potentially allowing for quick intervention. Such clinical interventions may also be enhanced and further understood through the use of this technology: While preliminary, research utilizing fMRI as direct assessment of drug action have become increasingly common and proves to inform dose-ranging investigations and enhance pharmacokinetic data (Matthews, 2006).

Functional decoding is another pioneering application of fMRI technology, constituting a more direct form of mind-reading: Decoding allows for predictions of the perceptions of an individual using their functional brain data. In the visual cortex, for example, decoding enables researchers to decipher the visual experience of an individual (e.g., what people are seeing in their dreams (Horikawa, Tamaki, Miyawaki, & Kamitani, 2013); or determine whether or not someone is currently viewing a sentence that is ambiguous or not ambiguous (Mitchell et al., 2004). Such analysis is intensive, and generally performed using machine-learning classification of the functional neuroimaging data, especially multi-voxel pattern analysis (MVPA) to detect changes in activity patterns across chunks of brain. While traditional univariate forms of whole-brain analysis simply search for correlations between fMRI time course and a reference (control data), thus assuming no spatial relevance for more “active” voxels or regions (and thus covariance across surrounding voxels is not taken into account), MVPA allows for the simultaneous analysis of voxels (i.e., does not consider each area to be independent of its neighbors). This helps to extract the signal from patterns of activation where, for example, a given voxel might not be significantly responsive to a condition in question but still related to it (Norman, Polyn, Detre, & Haxby, 2006; Haxby, Connolly, & Gunupalli, 2014).

Classification techniques rely on pre-recorded datasets to “train” classifiers by providing a subset of activation patterns with labels that represent the experimental manipulation into a classification algorithm. This allows for the “learning” of a function that meaningfully associates activity patterns with experimental conditions (see Figure 12; Norman et al., 2006).



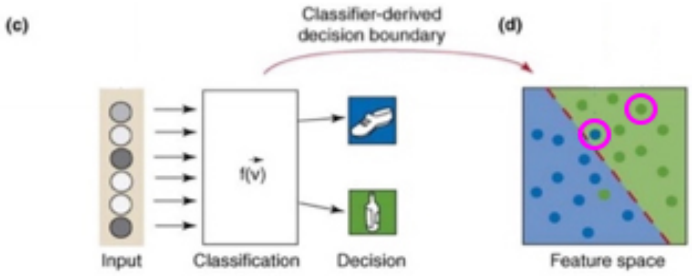


**Figure 12.** A representative model for how MVPA can be effectively utilized to classify brain states via fMRI time series data  a) Stimulus categories are chosen and feature selection identifies relevant brain areas for classification by analyzing which areas most highly represent the experimental manipulation. b) fMRI data patterns from relevant features at each time point in the series are labeled to correspond to their stimulus category (bottle vs. shoe). The pattern templates are separated into datasets that are to train the algorithm or later test the algorithm's performance. c) The algorithm (classification function; f(v)) is trained to identify patterns and their corresponding stimulus condition. d) The classifier-derived decision boundary is defined in multi-dimensional space (simplified into 2D here). Every data point on the feature space defines a pattern of activity over time and the color defines its experimental stimulus category, while the colors of the feature space defines the classifier's indication of category based on these patterns. The classifier algorithm can now predict stimulus category for the held-out test dataset: In this hypothetical example, the circled dot on the right represents a correct category classification (bottle-like pattern predicts bottle stimulus), while the dot circled on the left represents an incorrectly classified pattern (bottle-like pattern predicts bottle stimulus, even if the pattern was actually exhibited during the perception of a shoe stimulus). Adapted from "Beyond mind-reading: Multi-voxel pattern analysis of fMRI data," by Norman et al., 2006, *Trends in Cognitive Sciences, 10, 424-430*.

### *2.4 Real-time fMRI and Brain-computer Interfaces*

Having shown that algorithms can be trained to predict mental states, it follows that these predictions may be linked to virtual or physical commands. Implementations like this are referred to as *brain-computer interfaces* (BCI), which allow their user to control or communicate using algorithmic predictions of brain states based on brain data. BCIs are increasingly common (see Figure 13). They have been used recently, for example, to control neuroprostheses by imaging somatomotor BOLD activation patterns. This permits amputees or tetraplegics the ability to direct movement of artificial prosthetics using only neuroimaging technology and their willing of the limb to move (Lee et al., 2008) or move virtual cursors (Weiskopf et al., 2004). In most instances, the realization of these BCI applications also requires that the data be collected and analyzed at speeds such that the processing and implementation delay of neuro-generated features are imperceptible to the user of the device—coined to be *real-time* (RT). Lee and colleagues' (2008) demonstration of this similarly required the use of an "off-line" training session (without RT feedback) to define regions-of-interest (ROIs) in 3D activation maps (relative to resting baseline) that were exclusive to the imagining of one movement of the robotic arm. In the on-line sessions then, the BOLD signal generated by imagined movement in the ROIs signaled parameters to the robotic arm and allowed the BCI user to control the arm. Due to the movement-sensitive and magnetic nature of fMRI, the user was held in place with a head restrictor and visualized the movement of the robotic arm (based on BOLD contrast levels in motor ROIs) via MR-compatible visual goggles in RT. This allowed the modulation of activation in particular brain regions based on RT visual feedback. While these constrictions make this far from an ideal BCI prosthetic, advances in other imaging techniques, such as electroencephalography (EEG), offer mobile BCI paradigms.

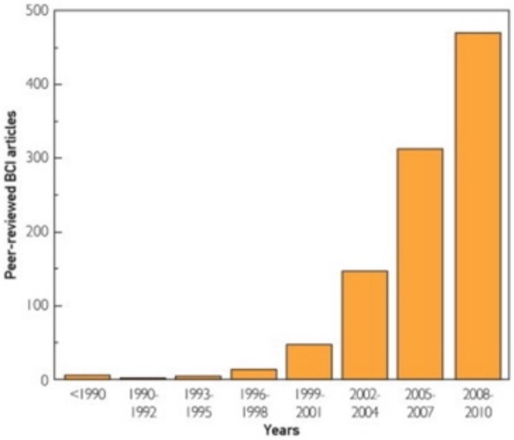


**Figure 13.** Peer-reviewed brain-computer interface publications have become increasingly common in the last 30 years. Reprinted from "Brain-computer interfaces in medicine," by Shih, Krusienski, & Wolpaw, 2012, *Mayo Clinic Proceedings*, 87(3), 268-279.

Evidence suggesting that individuals have the capacity to modulate activity in specific brain regions based on meaningful neurofeedback has spurred the imagination of neuroscience researchers. Growing evidence has also suggested the possibility of using BCI not just to alter our outside world, but to influence and augment the very cognitive assets we possess as well. The new frontier is no longer looking outward towards the horizon—it's looking inward.

For example, appropriate fMRI-based neurofeedback has shown to be able to target regions of the brain which are responsive to positive affective images in patients with major depression. The BCI users then successfully trained themselves to increase target area activity. Relative to non-neurofeedback controls, those in the active condition had significant improvements on measures of depression and also showed an increase of activity in the ventral striatum—which is notably involved in reward prediction and the development of reward-based behaviors, such as drug addiction and drug-seeking (Haber, 2011)—and other cognitive control regions throughout the training period (Linden & Lancaster, 2011). The treatment of disordered behavior and thinking with neurofeedback is an emerging avenue of clinical neuroscience. While the current results are fairly mixed (Fielenbach et al., 2018), such organic methods of brain state changes promise to give patients a sense of control over their experience of mental illness which might be lacking, especially for those with treatment-resistant symptoms or those who dislike—or have strong side-effects from—traditional medication interventions.

Recent attempts to implement non-clinical cognition-enhancing neurofeedback paradigms have also been met with some success. One such investigation (deBettencourt, Cohen, Lee, Norman, & Turk-Browne, 2015) used a type of BCI that allows for *closed-loop* neurofeedback. This approach utilizes MVPA-processed neuroimaging data to update task stimuli in a way such that the brain state triggered by a stimulus at one particular moment influences the presentation of the stimulus the next, which then changes the brain state as it reacts to the new stimulus, which triggers the stimulus to update, ad infinitum (hence *closed-loop*). In this study, deBettencourt and colleagues asked participants to selectively attend overlapping stimuli (i.e., 50% an image of face, 50% an image of a place), training a whole-brain RT fMRI classifier to discriminate between whether its user was attending one category or the other. On neurofeedback trials, then, the researchers altered the ratio of the task-relevant and task irrelevant stimulus on the screen based on the extent to which their detected brain state indicated that the participant was attending the task-relevant category of stimulus (see

Figure 14). This paradigm rewards task-relevant brain states by increasing task-relevant overlay and likewise penalizes task-irrelevant brain states, functionally magnifying the consequences of one's attentional state on the following trial.

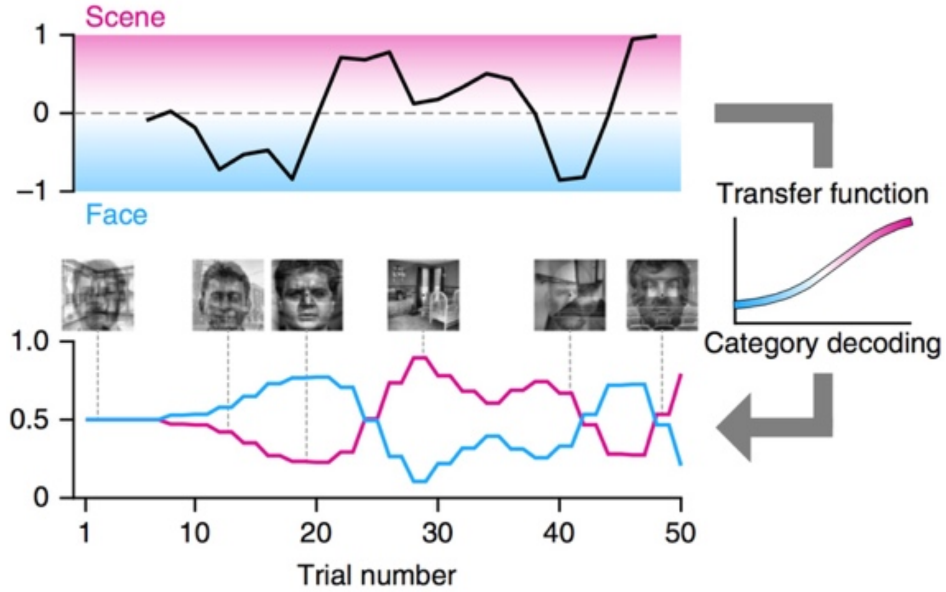


**Figure 14.** Category decoding (top) output informed the extent to which the participant was attending the overlaid face or scene stimuli. This prediction altered the ratio of face to place on the subsequent stimulus via a sigmoidal function (right) such that more attention paid to one category of overlay increased its stimulus proportion on the following presentation (bottom). Reprinted from "Closed-loop Training of Attention with Real-Time Brain Imaging" by deBettencourt et al., 2015, *Nature Neuroscience*, 18(3), 470-478.

Participants in this study were instructed not only to attend the task-relevant category of image and respond with a button press at each trial, but were also asked to withhold responses to a subcategory of task-relevant stimuli which occurred less frequently than the primary subcategory (i.e., a go/no-go task). For example, if the task-relevant category was places, they might be asked to withhold responses to indoor places. Behavioral performance, measured by the number of correctly responded-to or withheld responses, improved after one training session and RT neurofeedback relative to control participants who received sham neurofeedback (which corresponded to pre-recorded neuroimaging data from others' brains). The improvement was greatest when feedback carried information relevant to the frontoparietal attention network. Together, the results of this study indicate that failures of attention, and potentially other cognitive features, are not necessarily resultant from reaching a ceiling of capacity, but can be trained with neurofeedback.

In these studies, and others (e.g., Yoo et al., 2012), fMRI provided meaningful RT neurofeedback to participants. But in some paradigms, the caveats of fMRI might outweigh its benefits. A major caveat of fMRI is that the size of the datasets makes them demanding to manage or analyze, and difficult to do so without significant delay. The additional delay introduced by the flow of blood relative to neuronal activity ("hemodynamic lag") also constitutes a significant problem for real-time analysis. Furthermore, fMRI machines' innate lack of portability (regularly weighing in over 50,000 pounds) and exorbitant price tag (usually requiring at least a million-dollar investment) make the equipment prohibited to most. It has also been evidenced that much of the allure and perceived credibility of research using brain images is accounted for by readers' affinity for reductionist explanations (manifest in fMRI images) of abstract phenomena like cognition (McCabe & Castel, 2008). As such, fMRI may present challenges to some paradigms. However, in addition to fMRI, there has been an abundance of recent work applying BCI methodology to data from other devices measuring brain activity, for example, training classifiers of brain states for single trial EEG data.

While fMRI technology promises to reveal much about the functioning of the brain, its headline-generating glitz (McCabe & Castel, 2008) might overshadow the potential of another, slightly older technology currently more attuned for consumer BCI: electroencephalography (EEG).

## Section 3: EEG

### *3.1 Basis of the Electroencephalogram*

To understand how EEG works, it's necessary to go all the way down to the cellular level. Neurons are surrounded in a lipid bilayer membrane, a fatty barrier which allows it to regulate its internal contents from its external environment. By controlling the flow of electrically-charged ions ($Na^-$, $Cl^-$, $K^+$; the negatively charged ones are called anions and the positively charged ones are cations) through the neurons' membranes with selectively permeable ion channels and pumps—proteins that do exactly what it sounds like they do, pump ions across the lipid bilayer—neurons can change the electrical potential across their membranes. Membrane potential: The difference between the potential on the inside of the cell and the potential on the outside ($V_m = V_{in} - V_{out}$). The resting membrane potential of neurons—their default state—is about -60mV to -70mV (Holmes and Khazipov, 2007).

The brain consists of roughly a hundred billion neurons, all communicating by sending electrical signals along their synapses and exciting or inhibiting other neurons with these signals. In order to communicate, the neurons must generate a strong enough electrical current to activate other cells; this is called an action potential. These occur when the membrane potential reaches a threshold (around -55mV) that causes voltage-gated $Na^+$ (sodium ion) channels to open and sodium to rush into the cell, making the inside of the cell positively charged—depolarized—but only for about one millisecond before the $K^+$ (potassium) channels open and the positive charge is dispersed (Teplan, 2002; see Figure 15). The speed of this process makes it possible for neurons to "spike" rapidly—up to hundreds of time per second.

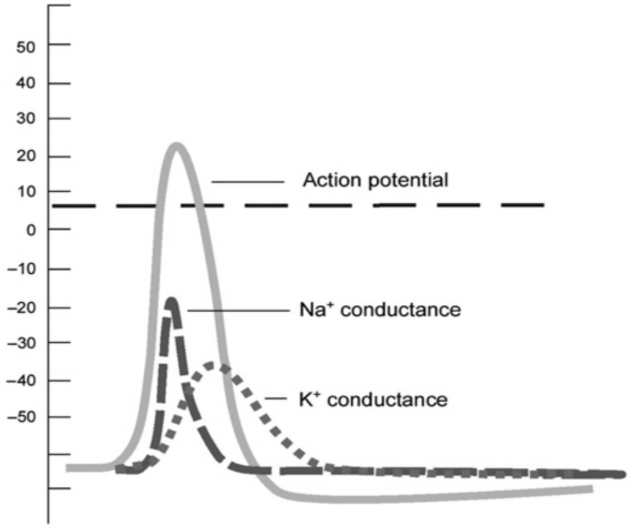


**Figure 15.** Opening of ion channels during an action potential. $Na^+$ rushes into the cell, making its charge more positive (increasing depolarization). This positivity causes more $Na^+$ to rush into the cell and depolarization accelerates, creating a positive feedback loop. Once sufficiently depolarized for the action potential to reach peak (creating an all-or-nothing effect), an influx of $K^+$ ions is triggered and $Na^+$ influx decreases; this results in an efflux of positive charge from the cell and it repolarizes back to the resting membrane potential. Reprinted from "Basic neurophysiology and the cortical basis of EEG" by Holmes and Khazipov, 2007, *The Clinical Neurophysiology Primer*, 19-33.

Extracellular electrodes can detect action potentials from individual neurons, but only if the electrodes are comparably sized (i.e. a few micrometers) and really close to the cell body. Conventional EEG electrodes—on the scalp and of a size on the order of centimeters—are thus much too far and much too large to detect these individual neuronal activities. So why introduce all of this information to talk about EEG? Well, neurons don't just fire alone. The action potentials they generate are sent down through the "axonal" connections (down the little wires called axons that transmit signals to other neurons' dendrites), depolarizing the membrane potentials in adjacent regions and causing other $Na^+$ channels to open; the firing of one neurons can result in the firing of others to which its connected, which then spreads from the originating neuron down through the others in a continuing cycle. When many neurons fire in unison, their summated potential (called a population spike) can be detected in EEG recordings. However, most of the signal detected by EEG actually results from post-synaptic potential—changes in membrane potential of the postsynaptic cell caused by neurotransmitters. Thus EEG is the measurement of electrical activity on the scalp resultant from synchronous activity of neural populations in the brain (Teplan, 2002).

The EEG signal being recorded is obviously, then, biased to the activity of closer, cortical neurons on the surface of the brain. But not only are EEGs biased to surface activity, they are also biased to specific orientations of neurons, namely the pyramidal cells that form a kind of columnar sheet on the cortex and provide much of the useful EEG signal (Holmes & Khazipov, 2007). So it would seem that deeper structures are not directly relevant to the signal; but because of the interconnected circuits of neurons in the brain, signals originating deep within the brain can have a substantial impact on the surface electrical activity (see Figure 16).

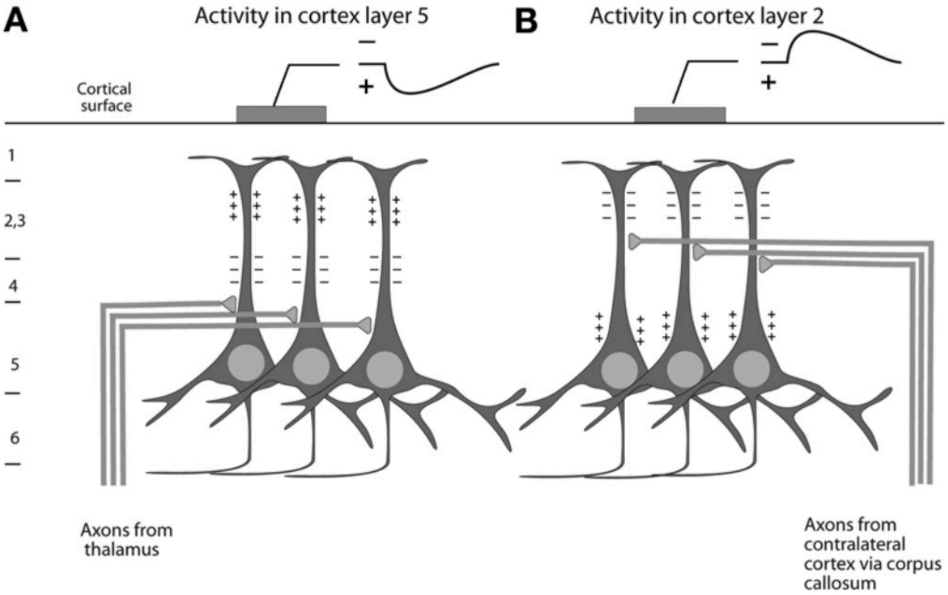


**Figure 16.** Surface EEG voltage patterns are dependent on the location of (even sub-) cortical activity. A) Input from the thalamic axons to these hypothetical pyramidal neurons would result in a positive-going surface potential while the input in B) would result in a negative-going EEG. Reprinted from “Basic neurophysiology and the cortical basis of EEG” by Holmes and Khazipov, 2007, *The Clinical Neurophysiology Primer*, 19-33.

Because this measure is generated by neuronal electrical potential, rather than relying on correlates of such activation—as in fMRI, which uses blood flow—the signal being recorded is a direct measure of electrical activity. EEGs can be obtained because the current generated in the brain flows between the electrical generator (source) and the recording electrodes on the scalp: Thus its major disadvantage over fMRI is that EEG provides a two-dimensional projection of a three-dimensional process; EEG cannot “find” the location of the source of the signals it records.

That being said, in its various forms, EEG has been used in creative ways to elucidate and explore inner operations of the brain. While its spatial resolution is extremely limited, as bias to higher-layer cortical neurons and signal interaction/cancellation entail (Tatum et al., 2008), EEG allows for excellent temporal resolution (1-10ms) in comparison to fMRI (Weiskopf et al., 2004). EEG and its devices are also much more accessible due to their lower price tag, cost of usage, safety, size, portability, and increased speed of data processing relative to fMRI.

### *3.2 Utilizing EEGs*

Brain patterns as detected by EEG form brainwaves that are fairly sinusoidal. Signals range from 0.5 to 100 µV in amplitude and are typically measured from peak-to-peak as a way to determine their underlying rhythmic frequency (Teplan, 2002). These signals also pick up subjects’ movements, even small ones like eye-blinks, and electrical noise present in the recording surround. Applying Fourier transformations to the EEG data can reveal the power spectra (rhythmic oscillations in voltage) of the data. While the power spectrum is technically continuous (ranging all the way from 0Hz to one half of the sampling frequency), today’s standard frequency bands are: delta, 0.5–4Hz; theta, 4–8Hz; alpha, 8–12Hz; beta, 12–30Hz; and gamma, > 30Hz. This delineation was done without much knowledge of the mechanisms at play and separated somewhat arbitrarily, though they have some relevance and drive much investigation still today (see Figure 17b). Thus, oscillations generated by the same mechanisms at different ages, across species, or even between individuals often fall into different bands without regard for their mechanics (Buzsáki, 2009). Nonetheless, brain states make EEG oscillatory frequencies more or less dominant over given time windows, and when data are averaged and combined at the group level, strong predictions can be made.

With contemporary EEG, there are three principal categories of features used to analyze human whole-brain electrophysiology: Event-related potentials, power spectra, and topography. The first technique is the evaluation of electrophysiological changes across the scalp time-locked to some event (often the onset of a stimulus), called *event-related potentials* (ERPs; Sutton, Tueting, Zubin & John, 1967). These are, in theory, the brain's direct reaction to a specific event, which cause short bursts of activity (called deflections) in response. Thus ERP deflections are typically very small voltage fluctuations only examined <1500 ms after the event (e.g., a stimulus presentation or a button-press) occurs and baselined to pre-event activity. The second common form of EEG analysis is the evaluation of frequencies, which describe the oscillatory patterns of activity across different areas. The third involves analyzing the topography of voltages or voltage patterns, or where the signals "travel" across a representational map of a scalp (e.g., Ko, Komarov, Hairston, Jung, & Lin, 2017). It's important to note that these features may be utilized in tandem for a given investigation and are not mutually exclusive. For example, a particular ERP deflection difference between conditions might only appear over a specific region of the scalp. See Figure 17 for a diagram of how these features differ and appear.

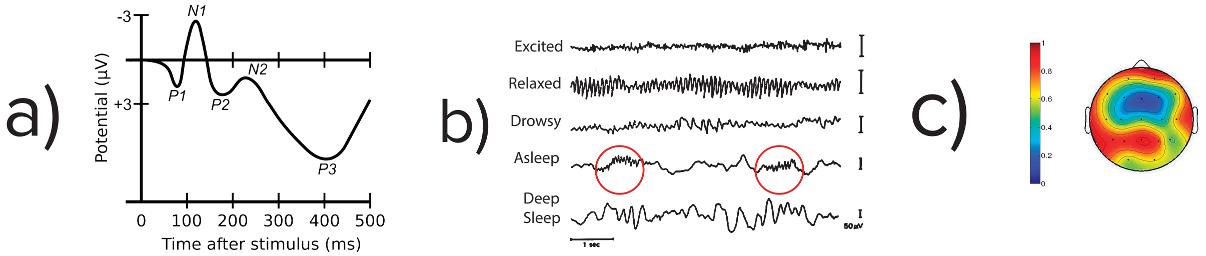


**Figure 17.** The forms that EEG data take. a) A representative ERP deflection. ERPs measure small electrophysiological changes (potential) after an event, typically a stimulus presentation. The P1, N1, P2, N2, and P3 indicators denote *components* of ERPs typically observed. b) Oscillations in electrophysiology were first primarily used (and still are) to quantify states of arousal and consciousness. An excited state is typically evidenced by beta waves, a relaxed state by alpha waves, a drowsy state by theta waves, sleep by the presence of sleep spindles (bursts of 12-14Hz activity, shown above in circles), and deep sleep by delta waves. c) A simple topographical map showing normalized amplitude over a time window. This might represent the magnitude of voltage, or the presence of a certain power band, over the scalp during an epoch of an experiment.

Due to the fact that EEGs are directly related to cortical activity, EEG analysis can assist researchers in determining and delineating the mechanisms facilitating particular cognitive functions or other brain states (e.g., Headley & Paré, 2017). For example, previous investigation has evidenced a robust involvement of EEG oscillations in attentional engagement: increases in frontomedial theta activity are often observed when sustaining attention; localized gamma oscillations promote activation of task-relevant processes across the brain; and the generation of alpha oscillations in task-irrelevant cortical regions decreases activity so as to inhibit distracting brain processes (Clayton, Yeung, & Cohen Kadosh, 2015).

In the memory domain, EEG studies can assist researchers in identifying the mechanisms of memory maintenance, encoding, consolidation, and retrieval (Düzel, Penny, & Burgess, 2010). A common means of identifying correlates of memory processes such as encoding is to compare data from when a participant is experiencing a stimulus that they later remember to data from when that participant experiences a stimulus they later forget: A difference in ERPs can be predictive of later memory (Paller, Kutas, & Mayes, 1987) and differences in oscillatory EEG patterns have also been shown to predict later memory (e.g., Hanslmayr, Staudigl, & Fillner, 2012)—these are called subsequent memory effects.

Applications of EEG technologies include rehabilitation, diagnostics (by identifying abnormal patterns), neuroentertainment (e.g., Yan et al., 2016, who triggered theatrical events on a stage based on the audience's level of engagement as determined via EEG; Dale, 2014), and cognitive training, many of which may seem quite distant from current possibility, but are closer than one might anticipate. For example, researchers can already use real-time EEG patterns to help control automated wheelchairs or regain motor abilities via neuroprosthetic limbs by having users simply imagine or intend the movement of that limb as well (e.g., Fok et al., 2011). These applications largely require the use of BCIs. Consequently, EEG BCI might sound like a natural step for this technology, but implementing EEG BCI in RT effectively brings many challenges; chiefly, data cannot be averaged over many trials, but need be processed on a moment-to-moment basis, and thus with a very small amount of data. Oscillatory patterns may have some basic advantages over ERPs in BCI for this reason; the small voltage differences present in ERPs typically require many trials to be averaged before meaningful information may often be detected (with the notable exception of some P3-based ERP BCIs, such a P3 spellers; see Fazel-Rezai at al., 2012 for a review of current applications and directions). while oscillatory investigations still require chunks of time before analysis can proceed, classifying the moment-to-moment activity by oscillations has shown to generally be more reliable. An additional challenge in EEG BCI is that it requires, unlike typical EEG procedures, that the pipeline for analysis be designed and implemented prior to data collection. For a typical EEG-based BCI neurofeedback pipeline, see Figure 18.

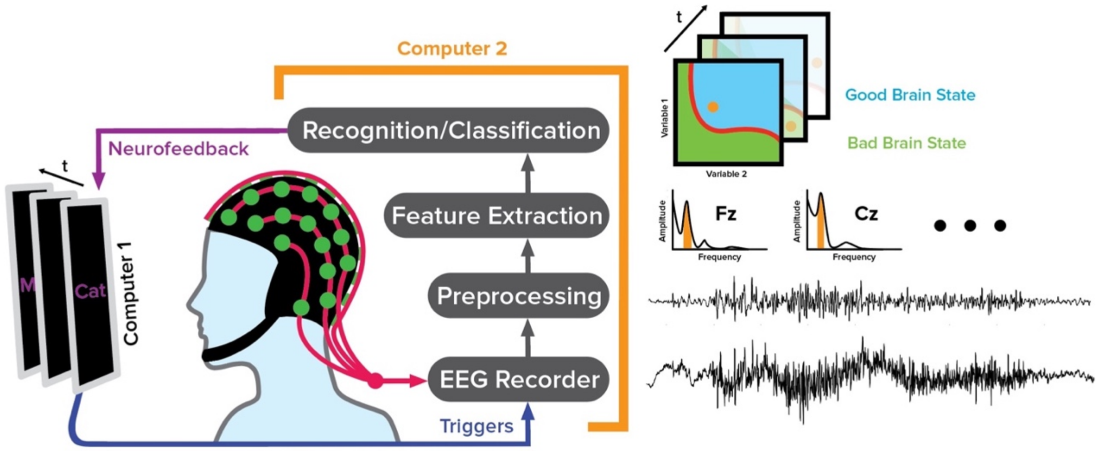


**Figure 18.** An example real-time electroencephalographic neurofeedback pipeline. Computer 1 sends trigger signals to computer 2 indicating where events are occurring in the time-series data being acquired by computer 2 from an EEG. The data are preprocessed, relevant EEG features are extracted from the dataset, and then chunks of data (relevant chunks indicated by the triggers) are classified based on a series of pre-established variable criteria. Based on the classification of the brain state data, neurofeedback is sent to the user or used to control prostheses/robots/software/etc. In this example, a good brain state permits the presentation of the stimulus (word on the screen to be learned), while a bad brain state withholds the stimulus.

### *3.3 Changing the Brain with EEG Neurofeedback*

In the course of the last twenty years, many EEG neurofeedback paradigms have evidenced the exciting possibility of monitoring and modulating the brain. One of the most common techniques involves an operant conditioning approach, whereby users monitor their own levels of activity in a neurofunctional feature of interest (e.g., activity in a region or power band of interest) and voluntarily control this signal by adjusting their behavior accordingly.

It has been shown in ophthalmic surgical settings that operant neurofeedback—in this study, informed by alpha-theta ratio and mu rhythm increases (without concurrent increases in theta and high beta; mu = 9-11Hz over motor cortex)—prior to surgery improves surgical techniques and decrease the duration of surgeries (Ros et al., 2009). Reddy, Rajan, Bagavathula, and Kandavel (2009) reported a case study of a patient with traumatic brain injury that resulted in detriments to memory functions; over the course of a neurofeedback training procedure, the patient was able to increase theta activity and decrease alpha activity, as the neurofeedback had intended, simply by viewing an indicator of these two features; and these changes were associated with improvements in verbal and visual learning and memory. Berman and Frederick report (2009)—in a similar procedure, over 30 sessions of 30 minutes each—normalization of EEG patterns and significant improvements in executive function and memory in patients who have dementia. Pineda and colleagues (2008) showed that improvements in sustained attention and the Autism Treatment Evaluation Checklist can be induced in children with autism after several weeks of training—in this paradigm, alpha band-controlled (more specifically, mu-controlled) neurofeedback affected the progression of a computer-based videogame or video. EEG-mediated neurofeedback has also been shown to decrease recidivism rates of incarcerates in proportion to the number of sessions they receive (Quirk, 1995). While these operant methods are of increasing interest, they are limited in many ways. Perhaps more complex forms of EEG analysis can provide new ways of harnessing the power of neurofeedback paradigm.

For example, Fukuda and Woodman published (2015) one of the first reports of effective, single-trial classification of cognitive states with EEG; they then improved participants' memory for “poorly-studied” images (as defined by their classifier) by having them restudy items in follow-up. The authors began by identifying two robust EEG subsequent memory effects from the literature that predict later memory: sustained positivity of ERPs in frontal electrodes and alpha activity at occipital sites at encoding. Using a simple behavioral paradigm in which participants studied images and then performed a confidence-rated recognition test for each image, the authors first confirmed the presence of these features in their EEGs and identified the time windows of greatest EEG differences between remembered and forgotten items (ERPs from 200-1000ms; occipital alpha from 400-1250ms). Following this, Fukuda and Woodman designed an online implementation of this framework in which a classifier reliably selected which items were more poorly studied using these pre-defined neural measures and targeted these items for additional study. The difference in baseline (prior to the neurofeedback stage) performance between well-studied items—categorized as being in the largest 40% of frontal positivity measures and in the lowest 40% of occipital alpha power—and poorly-studied items—in the smallest 40% of all frontal positivity and highest 40% of occipital alpha power—was eliminated after their re-study paradigm, which re-presented these images to participants. Studying the poorly-studied items (as defined by EEG measures) again had a 30% larger benefit of restudying than the benefit for restudying those items which were well-studied, suggesting that the detriments of natural encoding ability fluctuations can be minimized via relevant neurofeedback.

The use of these more modern computational methods has provided a robust avenue for future research in the field—and not just research, but also application. Mu, Hu, & Min (2017) have shown the potential of classifying fatigue level states of drivers via EEG, reporting over 98% classification accuracy across participants using entropy-based feature extraction (which measures the uncertainty or amount of information provided by an event) and a type of machine learning algorithm called a support vector machine (see Min, Wang, & Hu, 2017 as well). Support vector machines are classifiers which operate by mapping the data input into a higher dimensional (parameter) space and, subsequently, applying a hyperplane to the feature space which best separates the data classes ($f(x)=wx+b$, with w being the weight vector and b being the bias). The efficacy of this hyperplane is maximized by calculating the smallest margins from the hyperplane to the nearest data point (Güler & Ubeyli, 2007).

Furthermore, Huang et al. (2016) implemented an online, closed-loop, EEG-based fatigue detector which monitored oscillatory activity. These data were decomposed using independent components analysis and then a Fast Fourier Transformation was applied to derive the time-frequency patterns. Subsequently, the researchers provided auditory warnings to drivers when fatigue was detected in a virtual reality driving simulator; their results showed decreased reaction times and better detection of potentially dangerous driving situations when receiving neurofeedback.

Machine learning tools applied to BCIs often take a different approach than the operant paradigms introduced at the beginning of this sub-section. Instead of selecting features a priori which align with mechanistic (or potentially pseudo-mechanistic) predictions, the majority of these these approaches rely on training datasets which are acquired when a user or set of users performs a particular cognitive action. Instead of training the user to adapt to the structure of a neurofeedback paradigm (e.g., modulate alpha levels), by training a classifier on the data from pre-recorded sets, the feedback system adapts itself to the specificities of the users' brain—sometimes referred to as non-stationary signals (Krusienski et al., 2011). Typical methods of data analysis in neuroscience can be blind to the wealth of dynamics and variability when analyzing the "average" brain patterns; but the drawback to more adaptive styles with BCI is that it requires computational methods to extract complex and high-dimensional features in order to classify robustly.

### *3.4 Current Proposal*

One major challenge in EEG BCI efficiency is that many researchers select features of interest via the already existing categories of oscillatory activity. It has been suggested (Sherry & Schacter, 1987) that diverse demands throughout humans' evolutionary history prompted incompatible computational requirements of the memory and attention system and forced it to dissociate and specialize, but these brain regions are thought to still share many low-level features, including oscillatory activity—which is used to promote inter-regional communication and unify processing goals (Headley & Paré, 2017). As a result, it might seem reasonable that selecting these as features of interest is always valuable, but for that to be the case, the literature would have to maintain unified understanding of the functions of this communicatory pathway. However, the brain's oscillatory dynamics are largely elusive, and comprehensive accounts are lacking and attempts to simplify or ignore these complexities in EEG BCI often leave something to be desired.

For example, a recent investigation conducted by Salari & Rose (2015) used a trigger-based neurofeedback design in which participants were presented with novel information at the detection of high beta activity; they showed a clear memory-related advantage for processing stimuli in a brain state producing this power band—this particular investigation could not replicate the same findings with theta activity. Their theory-driven approach of using beta and theta activity was based on research indicating the involvement of these bands in memory encoding (e.g., Sholz, Schneider, & Rose, 2017). But because the mechanisms of oscillatory memory patterns are not fully explained, the simplicity of these more theory-driven methods leaves many questions unanswered, and might not demonstrate the full potential of these BCI systems for cognition-enhancing neurofeedback. In dealing with extremely dynamic processes like memory and attention, harnessing more flexible EEG-based cognition-enhancing BCI may prove to increase the strength and explanatory power of results.

Another major challenge in BCI applications is related to efficacy: the large inter-subject variability in the spatiotemporal characteristics of EEG. Prompting participants to adapt their patterns of brain activity to fit the requirements of a BCI system such that they can regulate control over specific EEG features (e.g., Birbaumer et al., 2000; Ros et al., 2009; Reddy et al., 2009; Berman et al., 2009; Salari & Rose, 2015) does not take into account the extent of this variability. Furthermore, even in clinical EEG neurofeedback studies which observed opposite effects in their physiological measures from prediction often report clinical improvements over non- or placebo-neurofeedback controls, demonstrating the wide variance in individual brain responses to neurofeedback (Fielenbach et al., 2018; Arns, Drinkenburg, Leon Kenemans, 2012). More adaptive BCI procedures (such as the Berlin BCI; Blankertz et al., 2007) allow greater control with minimal to no training intervention by identifying features in a subject-specific way using more flexible analysis like machine learning classifiers.

The goal of machine learning is to have computers make judgments about data without explicitly defining how it should be done. The computer learns from "experience"—training data provided which indicates the class the data belong to (e.g., epochs of EEG requiring top-down vs. bottom-up attention)—and develops a method based on this training data which "decides" which features are most relevant to or different between the classes. This is done offline by having a single, labelled dataset and performing crossvalidation, which essentially pulls out an epoch of data, runs the analysis on the rest of the data, and attempts to classify the pulled-out epoch based on the result of this analysis. So, provided new data (e.g., in real-time; "online"), that same classifier could perform computational operations on it with the “knowledge” of how to separate it into one class or another, and output a judgement. This is the basis of flexible paradigms, which prove useful in EEG BCI due to the low signal-to-noise ratio of (especially single-trial) EEG data, allowing for small (but potentially robust) and subject-specific class differences to drive the algorithms’ results.

The limited understanding of the direct link between mnemonic and attentional mechanisms and EEG signals is problematic for the development of a learning-enhancing EEG BCI. But it may perhaps be expedited by taking more data-driven approaches and then comparing results to contemporary theories as substantiation, rather than the other way around.

More flexible systems would also allow for learning improvements in sub-populations that may have limited access to or control over specific cognitive and neural factors that are conducive to learning. While much attention has been paid to improving learning capabilities in neurotypical people, specific individuals, or even the relevant sub-populations, few of these systems developed would be compatible across all groups if using rigid controls. Any particular individual might have a learning impairment specific to them and only them, thus systems designed to improve learning based simply on patterns of the general population might not be helpful. For example, someone’s memory abilities might be above-average, but their attention wanes might have a more extensive effect on their later remembrance than others; in data-land, this might manifest in a particular patterns of activity, thus an ideal neurofeedback paradigm would recognize this factor and adjust its predictions accordingly. Creating a system of neurofeedback trained on one’s own specific behavioral and neural data promises to facilitate a technique agnostic to differences between groups of people and individuals. Given the vast variations in cognition, brains, and the mapping of these two systems onto one another, the apparatuses of cognitive modulation that society chooses to invest in must be flexible in order to adequately address the whole of the population. The sciences, including applied neurosciences, have an oft-ignored responsibility to make its systems as all-encompassing of the diversity of its users as that diversity itself extends. Such systems also facilitate more functional apparatuses in general—in that if fewer people can use the apparatus, it necessarily falls short of its broader goals.

Lastly, while much of the psychological and neuroscientific research endeavors to understand the dynamics of learning focus on retroactively ascertaining the mechanisms and associations of better or worse learning, an ideal way of implementing such attempts would be to do so “at the gate”—before the effects take place. Think back to this paper’s very first example of a student in the library late at night: A person might push through the night trying to learn something for the next day, so telling them in the morning whether or not they did a good job of learning isn’t that useful of a feature. While much of the past research using brain data devices have examined relationships between set variables retroactively, an ideal system would harness these new technologies and implement behavior-modifying feedback in real-time and accounting for the users’ ability in the moment.

The current study draws on the literature and framework introduced herein to investigate the building of a BCI that encourages its users to adopt brain states that facilitate better learning. To do so, EEG data was recorded while participants completed a novel task developed for this project—the Attention/Memory Search-Manipulation and Recognition Task (amSMART). These data were then analyzed offline using EEG processing and classification techniques to determine how predictive and robust the data are for use in a neurofeedback paradigm. Subsequently, this could allow the presentation of neurofeedback to participants about the extent to which their current brain state is conducive to learning while they study novel information.

## Methods

The methods described in this section have been approved by the Bard College Institutional Review Board (see Appendix F)

### *Participants*

All 6 participants (3 female, 1 non-identifying) in this study were undergraduates recruited from Bard College and the surrounding area via snowball sampling and advertising in the college's campus center. All participants were between the ages of 18 and 35, had normal (or corrected-to-normal) color vision, were native English speakers, and did not have diagnosed attention deficit disorders, learning disabilities, or other neurological disorders.

Four participants were excluded (not included in total participant number above): In the first case, the participant's hair proved incompatible with the Emotiv Epoc+ device (some hairstyles are difficult for EEG devices in general due to the nature of the electrode and the volume of hair that would be required to be underneath the electrode); in the second case, the EEG acquisition software provided from Emotiv crashed during the EEG data collection phase; in the third and fourth cases, EmotivPro failed to insert triggers into the recorded data. In instances of procedure failure, as soon as the error was realized, participants were debriefed, dismissed, and compensated like participants who completed the procedure fully.

### ***EEG materials***

Participants in this study were outfitted with an Emotiv Epoc+, an electrophysiological interface for consumer use—with potential for research, as well. See Badcock et al., 2015; Maskeliunas, Damasevicius, Martisius, & Vasilievas, 2016; Badcock et al., 2013; and Ekandem, Davis, Alvarez, James & Gilbert, 2012 for published research evaluating Emotiv's neuroheadsets. Furthermore, classifiers built using data from Emotiv's hardware has already been used to detect levels of memory competition and predict memory retention (Rafidi, Hulber, Pacheco, & Norman, forthcoming). The Epoc+ was used in place of traditional, higher-resolution alternatives for ease of use and to test how currently realizable cognition-enhancing neurofeedback is at the level of an everyday consumer. This device and its software also interfaces with a developing field of EEG software applications from third parties (such as FieldTrip, OpenVibe, Neuropype, who provide drivers and other software for real-time streaming BCI, etc.); thus this device was also chosen in order to expand the potential accessibility of the paradigm and the ability to port similar designs to new systems.

The Epoc+ device is a consumer EEG headset containing 16 wet electrodes aligned with the 10-20 EEG system (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, FC4, M1, and M2); left and right mastoids (M1 and M2) were used as ground and reference positions (M1 acts as a ground reference point for measuring the voltage of the other sensors, while M2 acts as a feed-forward reference point for reducing electrical interference from external sources), providing a baseline voltage comparator for the other 14 electrodes from which data are actually saved (though data were re-referenced prior to analysis, see Data Methods). Gyroscopic information is also gathered during acquisition but was not used in the current experiment. Channel mapping is depicted in Figure 19 and corresponds to the standard 10-20 EEG system. However, due to the fixed positions of the electrodes relative to the headband (each electrode is connected via flexible plastic arms), variability in head-shape may have greater effects on electrode placement than in traditional EEG. The headset is entirely wireless, relying on a lithium battery for power and transmitting the data over Bluetooth to connected devices via USB. The sampling rate of the Epoc+ is 128 Hz.

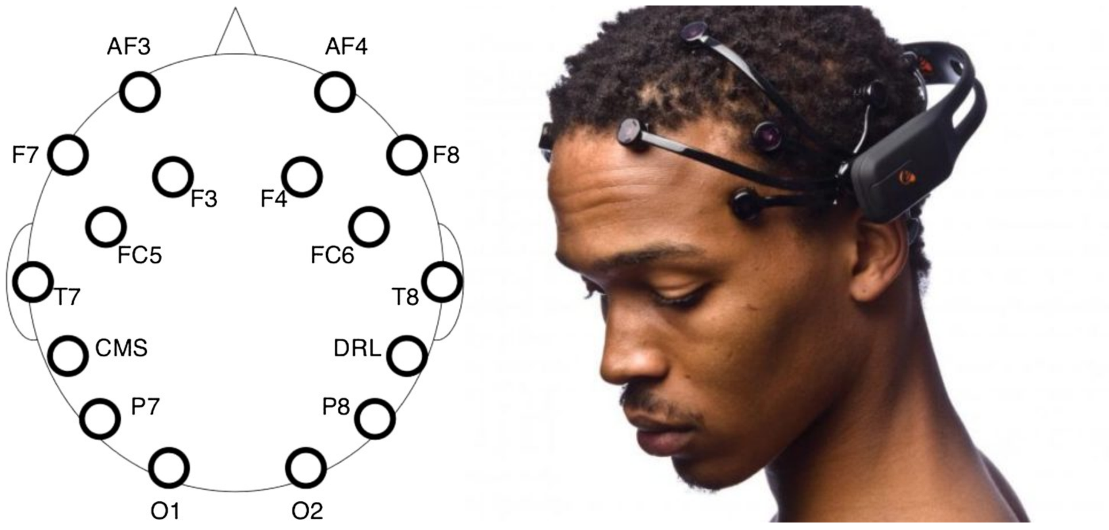


**Figure 19.** The Emotiv Epoc+ neuroheadset is composed of 14 electrodes and two references (here M1 is labeled CMS; M2 is labeled DRL) and mapped to the scalp according to the 10-20 standard layout. Reprinted from “Emotiv User Manual” by Emotiv.

While EEG data quality is crucial to successful BCI performance, traditional, gel-based EEG sensor rigs that offer higher-quality data are (1) much more expensive—easily exceeding 50,000 dollars for a medical-grade machine (for comparison, the Epoc+ costs about 800 dollars currently) (2) require much longer setup times than consumer devices such as the Epoc+ (traditional EEG capping takes >15 minutes, while the Epoc+ can take only a few seconds to establish a good signal) (3) involve more extensive clean-up processes which require careful cleaning of each sensor so as not to damage the electrodes and (4) another individual present to “cap” the EEG system (to put it in place on the scalp), while the Epoc+ is easily placed on the scalp by oneself. Thus EEG systems like the Emotiv Epoc+ that are portable, (relatively) cheap, and easier to use would be more accessible to the current consumer.

Electrode impedance was decreased by soaking proprietary felt pads purchased from Emotiv in a light saline solution. Emotiv’s custom system of signal quality detection (in EmotivPro) was used to gauge signal quality before the experiment; every electrode for all participants achieved Emotiv’s proprietary “green” signal, a measure of electrode impedence (in the 10-20 kΩ range; Duvinage et al., 2013) prior to data acquisition and were monitored throughout the experiment. Problematic electrodes (those that dropped from “green” to “orange” were re-wetted at the half-way point of the experiment and no electrode signal quality indicator dropped below “orange” at any point during data collection. Raw EEG data were acquired using Emotiv’s EmotivPro software.

### *Behavioral Materials: The amSMART Paradigm*

This preliminary research involved the development of a novel task, amSMaRT, which is split into two distinct phases.

The first phase consists of a computerized task (see Figure 20) in which participants complete two types of visual searches: A “difficult” visual search task and a relatively easier visual search task. In both types of visual search, the participants are instructed to find the red T (the target), among a 5x5 grid of Ts and Ls. Each trial of the task consisted of 25 Ts and Ls on the screen, overlaid on a black and white “distractor” image. In the easy visual search trials, the only red object on the screen is the target (feature search), while in the difficult visual search trials, each L was randomly assigned to be either blue or red (conjunction search), adding an additional source of complexity that requires greater attention to the task. Participants are instructed to press the key (1-5 on the number row) corresponding to the column in which the red T was present (e.g., in Figure 20, the participant would be tasked to press ‘2’).

Each image is presented on the screen for 8 seconds, during which time participants complete a series of visual search tasks of one type (i.e., a series of difficult searches or a series of easy searches). After the 8 seconds have elapsed, the image disappears from the screen, but the visual search display remains until a response to that array is registered. Once this response is registered, a new image and visual search display (equal chance of difficult and easy type) is presented; ad infinitum, until all images (144 in the current study) are presented.

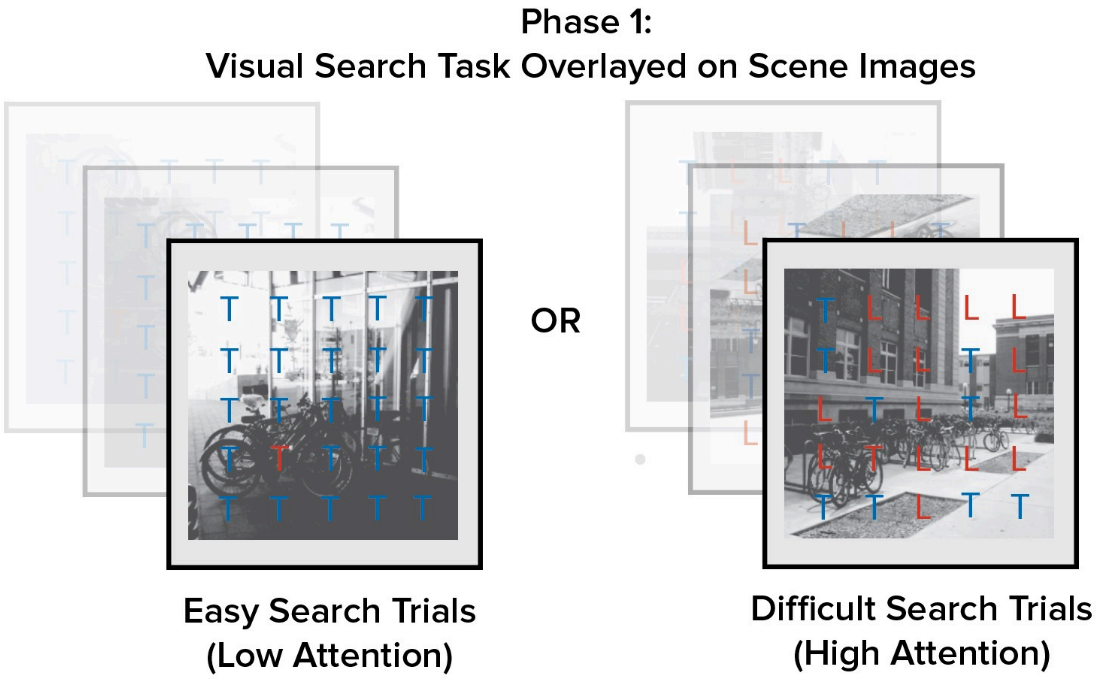


**Figure 20.** Phase 1 of the amSMART paradigm consists of either a series of “easy” visual search trials (left) or “difficult” visual search trials (right) overlaid on a grayscale scene image.

The second phase of amSMART consists of a 2-stage recognition test for the black and white scene images that have been previously presented underneath the visual search displays (see Figure 21). In the first stage, a binary decision is made between two images on the screen: One of the images has been presented previously and is the correct image to choose, while the other image is a lure—a picture from the same scene category (e.g., if the correct image was a picture of a blue airplane in flight, the lure might be a picture of a red airplane on tarmac). The decision is then weighted in the second stage based on the participants’ confidence in their previous response (see Data Methods for more detail). Both stages are repeated in Phase 2 for every image presented in Phase 1.

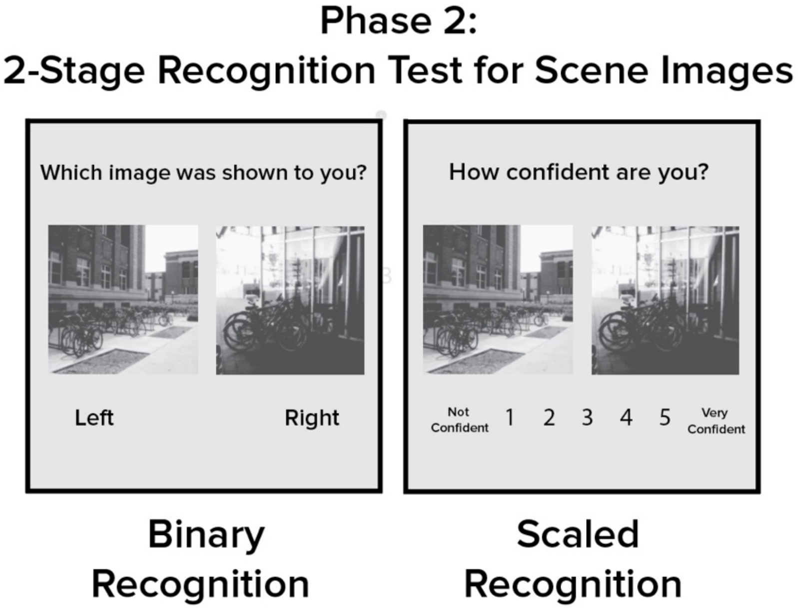


**Figure 21.** Phase 2 of the amSMART paradigm consists of a two-stage judgment for each scene image viewed in Phase 1. First, participants make a binary choice between the previously viewed image and a category lure image, then participants rate their confidence in that choice.

### *amSMART Procedure*

Participants first read the consent form (see Appendix A), and—upon providing consent—were outfitted with the Emotiv Epoc+ neuroheadset, sat, and placed their chin on a table-mounted chinrest at a fixed distance from the screen (visual angle approximately 15.5º). In consultation with the EmotivPro software display, electrodes were placed and replaced and hair was adjusted underneath the electrodes until signal quality achieved Emotiv's "green" signal across each electrode.

Participants were instructed that the task they were about to complete is a visual search task, which involves trying to locate a letter of a particular color among an array of other, colored letters; they were told that any image presented behind the visual search display was included as to distract the participant from paying attention to the primary task (the visual search). The researcher explained that keyboard responses should be made as quickly as possible without making mistakes. Participants were given four breaks during this phase, evenly distributed throughout the phase by portion of stimuli viewed.

At the completion of the first phase, participants were given Sudokus (see Appendix B) to complete for 5 minutes as a filler task between Phase 1 and 2 in order to reduce recency effects on Phase 2 memory data. Contact quality was also monitored throughout Phase 1 and electrodes were readjusted at the end of Phase 1 in the case of "imperfect" (according to the aforementioned proprietary Emotiv signal quality indicator) signal quality.

Before the second phase, participants were informed that the next section of the amSMaRT would be testing for the images that they'd previously considered to be distractors. After being instructed on how to respond to the display prompts, participants completed the two-stage recognition task for each scene stimulus.

After completion, the Epoc+ was removed, a post-experiment questionnaire (see Appendix D) was completed, and participants were debriefed and dismissed. Participants were compensated with entry into a raffle for gift cards ($50).

### *Apparatus*

This research was conducted using a Dell XPS13 running Ubuntu 16.04 for the task display, and a Dell XPS15 running Windows 10 for data acquisition, processing, and analysis. Stimulus presentation was performed using Matlab and Psychtoolbox-3. Data were collected using EmotivPro's acquisition service and analyzed using Matlab, Fieldtrip toolbox, and Donders Machine Learning Toolbox (DMLT). Event markers were sent over 2 male-to-female USB-serial port converters attached to a male-to-male serial cable from Psychtoolbox to EmotivPro, and a trigger output duration of .0076 seconds was selected for a single signal to be detected from the Psychtoolbox function by EmotivPro.

### *Data methods*

EEG data were loaded into Matlab and reorganized to match the data formats compatible with Fieldtrip. These data were then categorized into four discrete bins: data corresponding to performing high attention (difficult search) trials in which the background image was previously recognized *and* done so with a confidence rating higher than 1 (hAhM; high attention, high memory); data corresponding to performing high attention (difficult search) trials in which the background image was previously not recognized and done so with a confidence rating higher than 1 ( (hAlM; high attention, low memory); data corresponding to performing low attention (easy search) trials in which the background image was previously recognized and done so with a confidence rating higher than 1 ( (lAhM; low attention, high memory); and data corresponding to performing low attention (easy search) trials in which the background image was previously not recognized and done so with a confidence rating higher than 1 ( (lAlM; low attention, low memory); see Table 1 for a visual layout. These data were not segmented, epoched, or binned by Fieldtrip Toolbox due to differences in EmotivPro's output EDF+ format from traditional EDF+ formatting; instead, these operations were performed "manually" using Matlab.

| | |
|---|---|
| **hAhM (bin 1)**<br>• Participant was completing the difficult visual search task<br>• Participant indicated the correct image (confidence > 1) | **hAlM (bin 3)**<br>• Participant was completing the difficult visual search task<br>• Participant indicated the incorrect image (confidence > 1) |
| **lAhM (bin 2)**<br>• Participant was completing the easy visual search task<br>• Participant indicated the correct image (confidence > 1) | **lAlM bin (bin 4)**<br>• Participant was completing the easy visual search task<br>• Participant indicated the incorrect image (confidence > 1) |

**Table 1.** Data bins extracted from continuous EEG data according to the difficulty level of the visual search task during each image presentation in Phase 1 and the participants' subsequent responses to the image in Phase 2.

Emotiv's output also automatically implements built-in digital notch filters at 50 and 60 Hz, and subsequently a built-in band-pass filter of 0.2–45 Hz; no filters were otherwise applied to the data. A select few trials with extreme values, which typically reflect muscular or ocular artifacts, were rejected manually and blind to data condition using Fieldtrip's visual inspection GUI (average rejected per person was 2.86 trials). Data were preprocessed using Fieldtrip and re-referenced to every electrode, rather than Emotiv's default of referencing to the left mastoid sensor. A custom layout was also built based on the 10-20 system to fit the Emotiv electrode placement and input to Fieldtrip at this stage as well. The EEGs were then averaged separately and in two different ways for the purposes of data exploration and ascertaining the most efficacious processing pipeline: by "raw" voltages and by frequencies.

Timelocked averages were computed per trial based on Voltage (amplitude; channels x time) using Fieldtrip. Timelocked time-frequency representations were generated using sliding wavelet convolution (mtmconvol) with hanning tapers per frequency window of 1 Hz. Classification was performed using a Fieldtrip wrapper for DMLT functions. First, individual data was normalized (z-scores were calculated such that data have a mean 0 and standard deviation 1). Subsequently, a linear support vector machine was applied to the data. Behavioral data were analyzed and visualized in Matlab and Excel.

## RESULTS

Visual search errors (selection of non-target columns) were relatively consistent across blocks of the amSMART (see Figure R1), suggesting that participants engaged with Phase 1 fairly uniformly throughout the duration of the task.

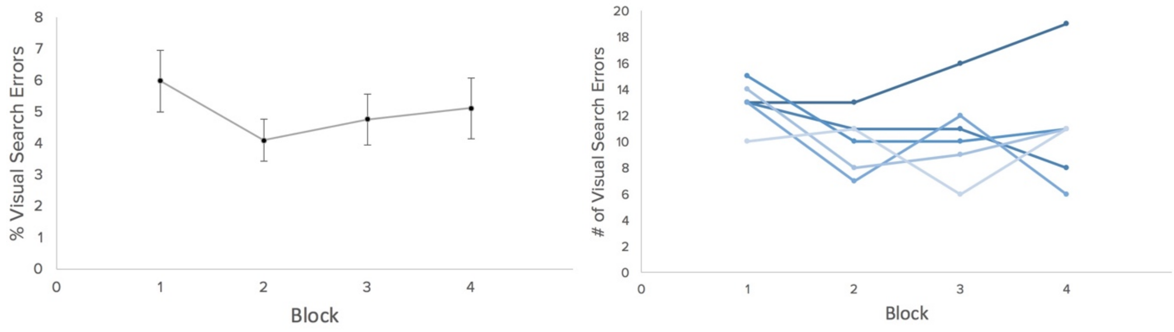


**Figure R1.** Left: Percent of incorrect responses registered across blocks of Phase 1 of the amSMART (error bars = SD). Right: The number of incorrect responses registered across blocks of Phase 1 of the amSMART by participant.

Moreover, visual search error rates did not differ drastically between participants or between difficult search task trials and easy visual search task trials (Figure R2). This indicates that the additional attentional burden in difficult trials did not have a substantial impact on visual search target selection accuracy.

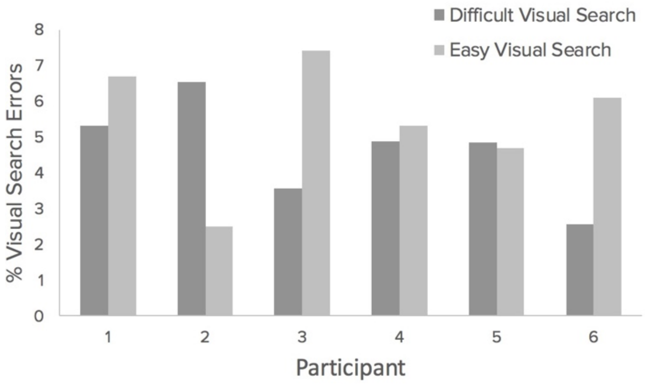


**Figure R2.** Percent of visual search errors by participant during difficult and easy visual search trials.

In Phase 2 of the amSMART, while rates of recognition remained relatively stable (Figure R3a), recognition confidence waned slightly throughout the phase (Figure R3b); concomitant was a decrease in reaction time throughout the duration of Phase 2 (Figure R3c).

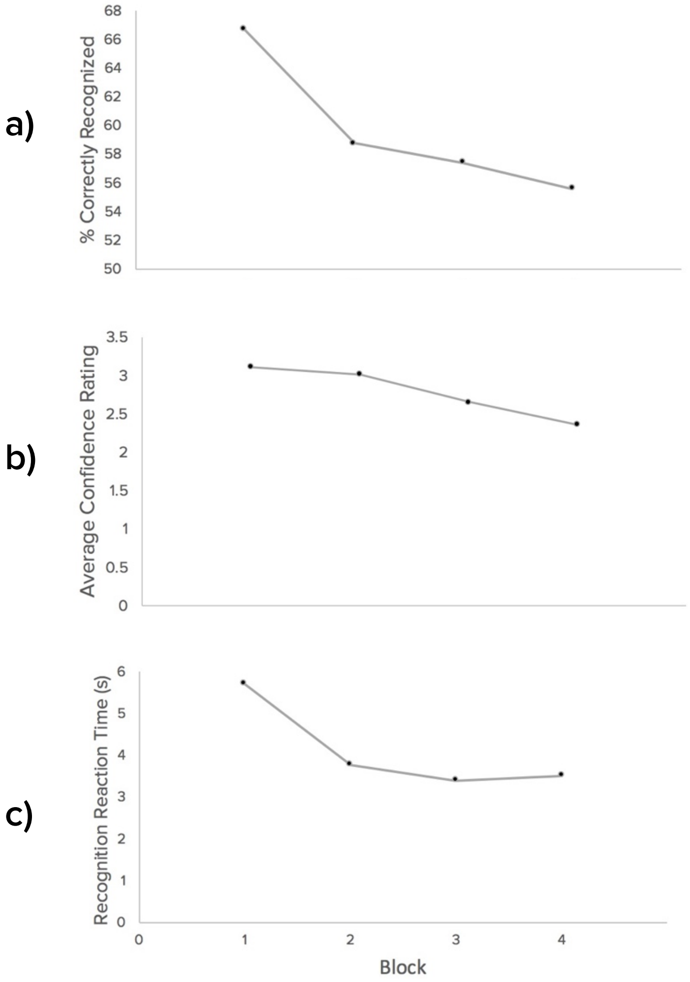


**Figure R3.** Averaged performance in Phase 2 of the amSMART across participants. **a)** Percentage of correctly recognized images per block. **b)** Average recognition confidence rating per block. **c)** Average recognition reaction times per block.

Recognition accuracy across Phase 2 of the amSMART appeared to increase with participants' subjective confidence ratings of their recognition judgments (Figure R4).

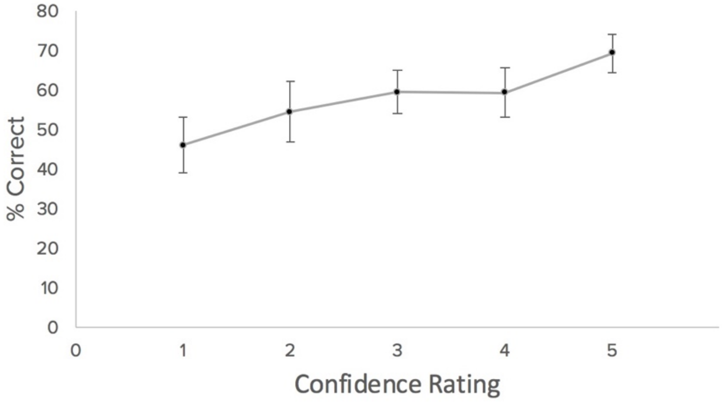


**Figure R4.** Percent correct recognition responses by subsequent confidence rating. Error bars = SD.

Binary classifier performance (either across remembered and forgotten item bins—Figure R5, top—or across bins requiring the same level of attention—Figure R5, bottom) was higher for every participant by memory than by attention.

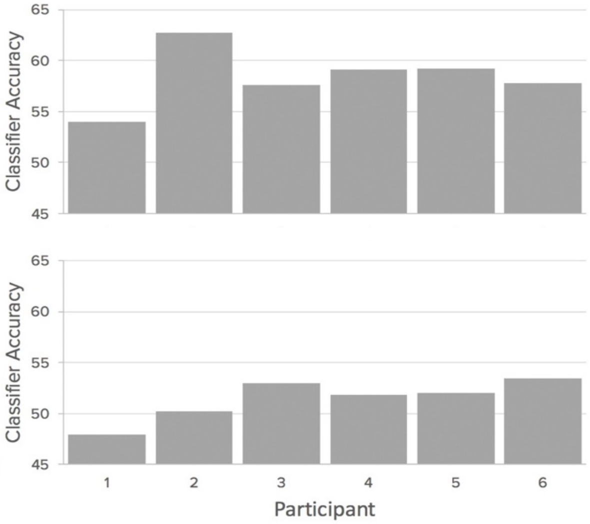


**Figure R5.** Classifier performance for memory (top) and attention (bottom) by participant.

Separating the data by mode of classification reveals further information about classifier performance (Figure R6). Classifying memory by frequency improved classification accuracy for all but one participant. The same trend was revealed in the attention classifiers, but even more robustly (Figure R7). In the current paradigm, classifying by frequency with the current method has an accuracy of about 55%; doing so by voltage revealed chance-level classification rates. Averaging across participants might make this effect clearer (see Figure R8).

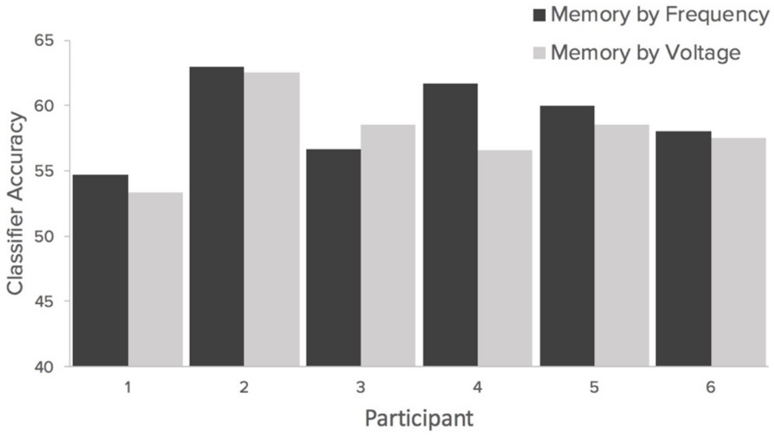


**Figure R6.** Accuracy improved for all participants except one when classifying across memory bins by frequency rather than by voltages.

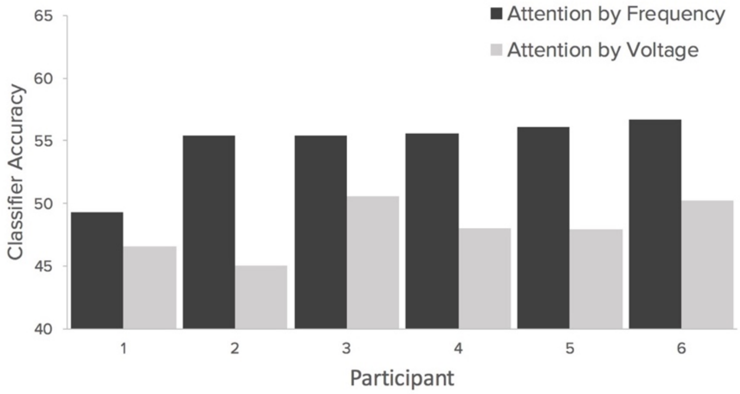


**Figure R7.** Accuracy improved for all participants when classifying across attention bins by frequency rather than by voltages.

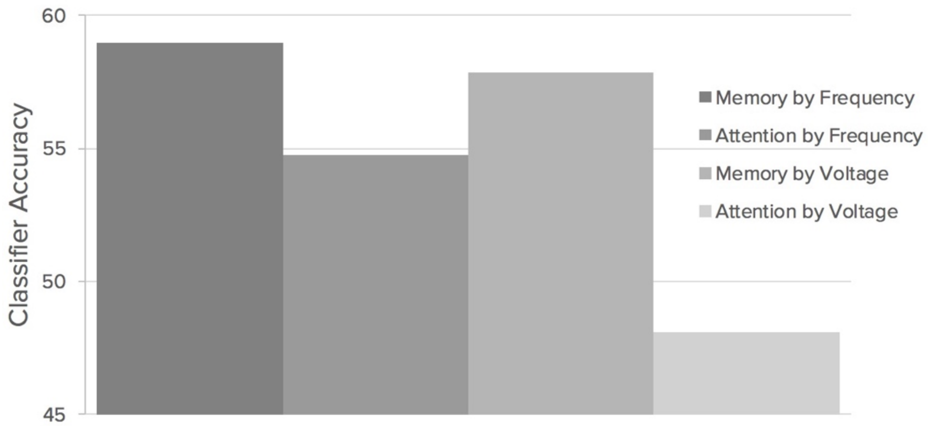


**Figure R8.** Accuracy improved across both attention bins and memory bins when classifying by frequency rather than by voltages.

Classification accuracy (across conditions *and* modes of analysis) seems relatively unaffected by the number of folds used in the crossvalidation procedure (Figure R9)—attention by frequency may see some benefit around the 15-fold mark (Figure R9b).

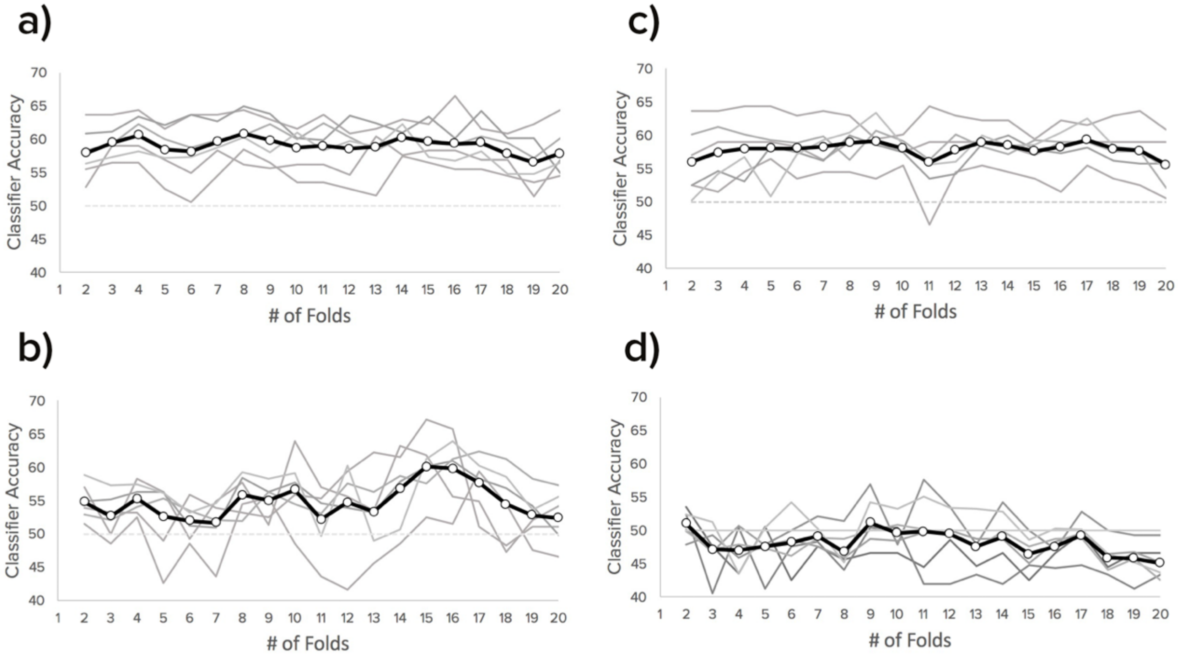


**Figure R9.** K-fold crossvalidation accuracy by k across attention bins (b, d) and memory bins (a, c), for voltage-based (c, d) and frequency-based (a, b) analyses.

Signals relevant to memory classification seem to be distributed throughout the amSMART trial, though slight advantages might be seen for the first half of the trials for memory (Figure R10).

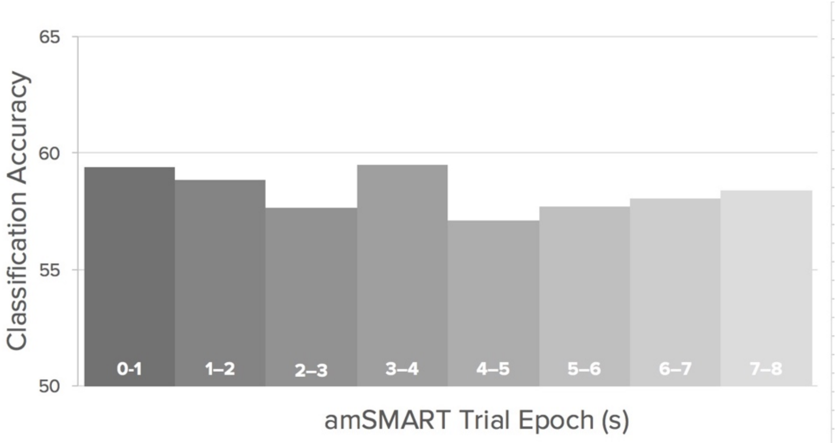


**Figure R10.** Classification accuracy for memory using only one-second epochs of the amSMART trials. Signals relevant to successful classification are distributed throughout the trials. All participants; all frequencies; 4-fold crossvalidation.

However, signals relevant to attention classification seem to be less distributed and more condensed towards the beginning of the amSMART trial (Figure R11).

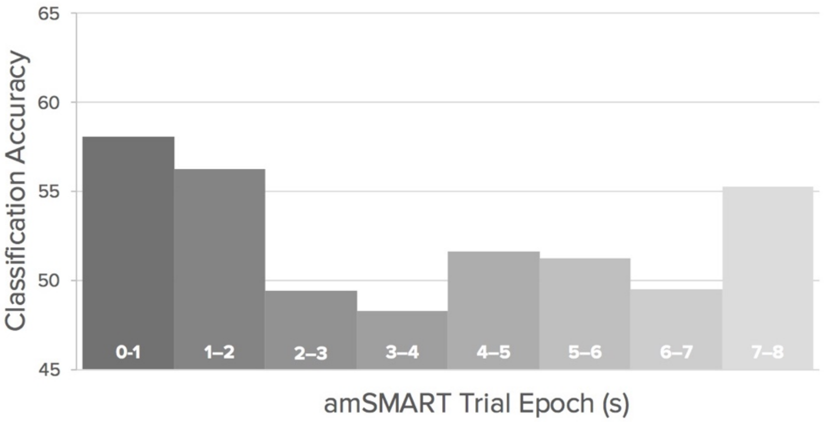


**Figure R11.** Classification accuracy for attention using only one-second epochs of the amSMART trials. Signals relevant to successful classification are mostly early in the trial period, though the "7-8" epoch may have relevance as well. All participants; all frequencies; 4-fold crossvalidation; note the different y-axis scale than Figure R10.

Signals relevant to memory classification also seem to be distributed throughout the different frequency bands, though slight advantages might be seen for analysis over theta and high beta activity (Figure R12).

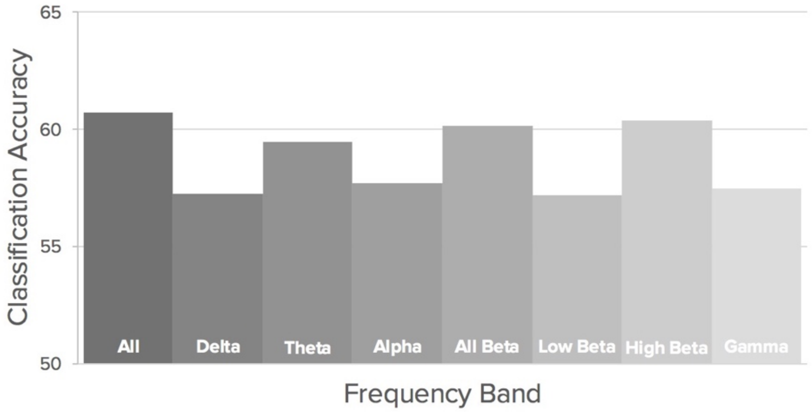


**Figure R12.** Classification accuracy for memory using only delta (.5-4Hz), theta (4-8Hz), alpha (8-13Hz), all beta (13-30Hz), low beta (13-18 Hz), high beta (18-30 Hz), and gamma (30-45 Hz) activity. Signals relevant to successful classification seem relatively distributed throughout the bands. All participants; 4-fold crossvalidation.

However, signals relevant to attention classification seem much less distributed throughout the different frequency bands; successful classification appears to rely heavily on theta and alpha activity. (Figure R13).

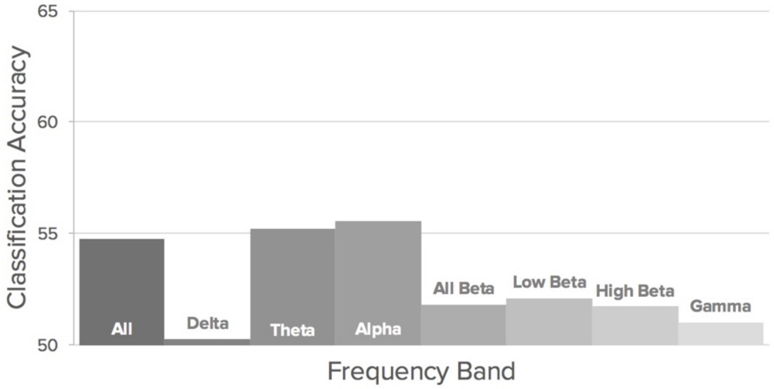


**Figure R13.** Classification accuracy for attention using only delta (.5-4Hz), theta (4-8Hz), alpha (8-13Hz), all beta (13-30Hz), low beta (13-18 Hz), high beta (18-30 Hz), and gamma (30-45 Hz) activity. Signals relevant to successful classification seem primarily to be present in theta and alpha bands. All participants; 4-fold crossvalidation.

Topographic plots of the output weight matrices reveal that, on the group level, there is some overlap between the signals being weighted for classification of attention (Figure R14 left, frequencies; Figure R15 left, voltage) and memory (Figure R14 right, frequencies; Figure R15 right, voltage). However, there are also some differences to be noted: Namely, occipital electrodes (particularly O2) seem to be weighted heavily for high attention classification, but the same electrode is weighted highly for low memory classification. It seems that there are some spatially distinct signals which may be reliable for classification.

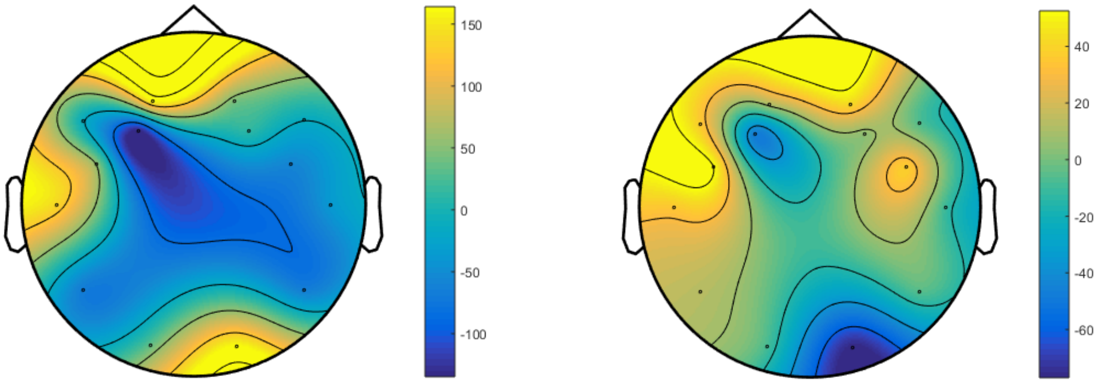


**Figure R14.** Topographic maps of classification weights across attention (left) and memory (right) bins by frequences. All participants; 4-fold crossvalidation; all 8 seconds of data; note the different colorbar scales.

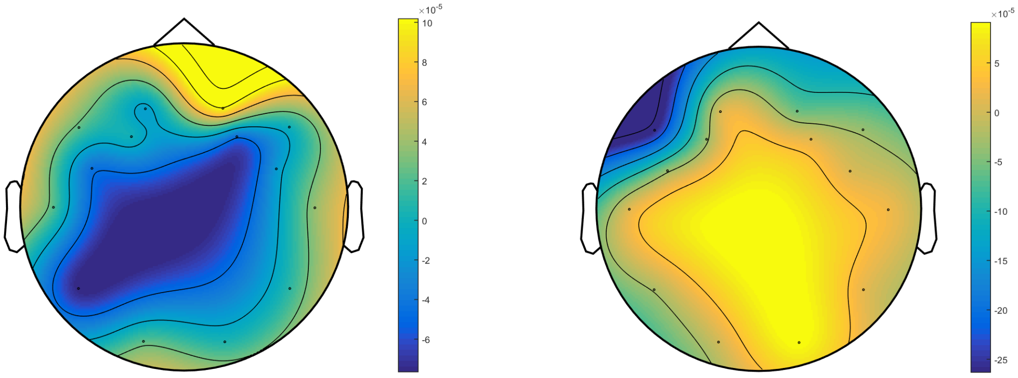


**Figure R15.** Topographic maps of classification weights across attention (left) and memory (right) bins by voltage. All participants; 4-fold crossvalidation; all 8 seconds of data; note the different colorbar scales.

However, these average weight values do not necessarily reflect the distribution of classifier weights of each participant. For example, there appears to be substantial variation between participants in the spatial distribution of signals useful for discriminating between attention classes (see Figure R16).

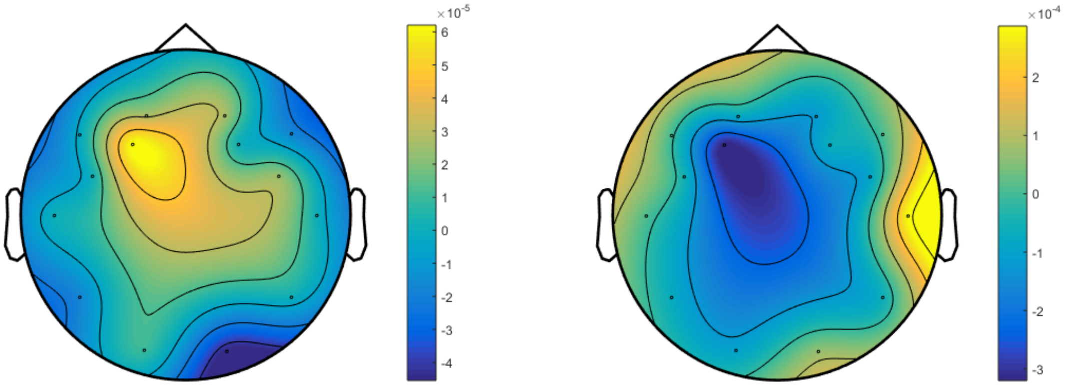


**Figure R16.** Topographic maps of classification weights across memory bins for participant 5 (left) and participant 6 (right). The spatial distribution of classifier weights varied considerably between participants. Voltage; 4-fold crossvalidation; all 8 seconds of data; note the different colorbar scales.

To further illustrate this point: Despite the fact that attentional state prediction using only gamma activity seems to be classified at chance-level (see Figure R13), doing this analysis per participant reveals that group-level activity appears to be a poor representation of the usefulness of this band in attentional state prediction. While four participants exhibit chance-level accuracy, two participants' attentional states were able to be classified with over 60% accuracy and significantly from chance level as indicated by a binomial test (Figure R17). Result such as this are typical of the exploration thus far.

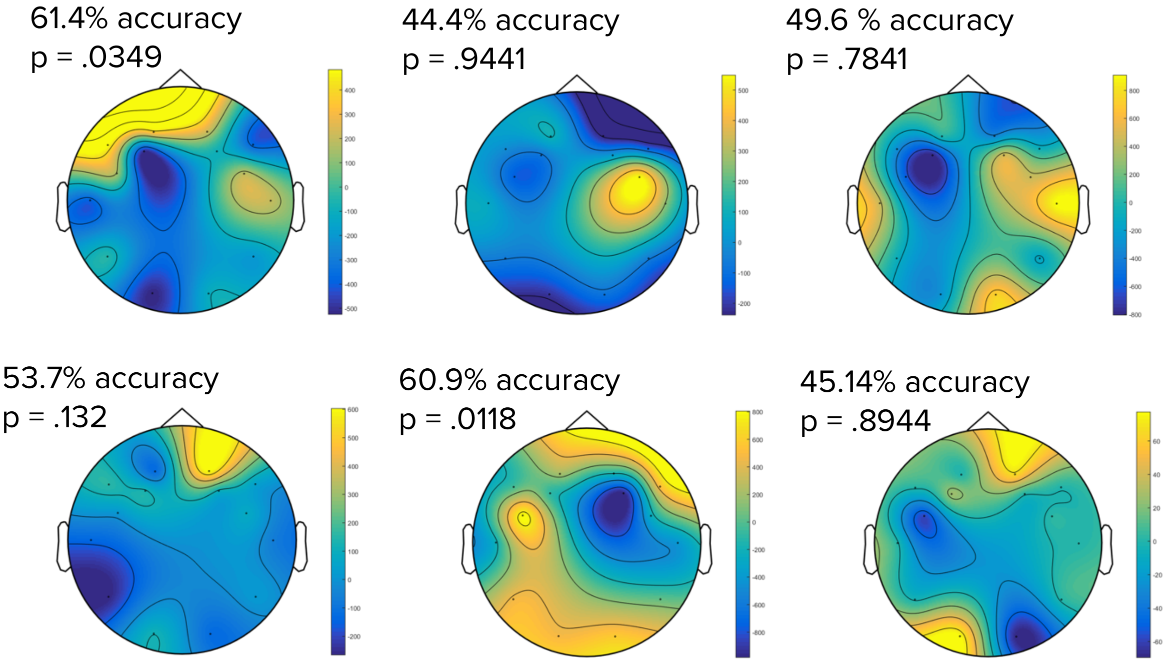


**Figure R17.** Topographic maps of classification weights across attention bins for participants 1–6. Gamma Frequencies (30-45 Hz); 4-fold crossvalidation; all 8 seconds of data; note the different colorbar scales.

# DISCUSSION

The ongoing, preliminary work presented here represents exploratory steps in the development of a neurofeedback paradigm for learning enhancement. After surveying the cognitive infrastructure for learning and the technical tools used to uncover and push the boundaries of the mechanics of learning, a novel task was tested for the purposes of ascertaining meaningful correlates of brain activity related to this infrastructure. Furthermore, this project aimed to determine whether EEGs recorded with consumer-grade hardware were viable replacements for higher resolution and established tools for acquiring brain data.

Overall, there appears to be a classifiable signal detectable with consumer-grade EEG across attention and memory conditions of the amSMART. However, classification performance across most conditions is low on the group level, and subject-level exploration indicates that each user has potentially distinct signals which may not generalize to other users' brain activity. Though classification performance, at its best, appears reasonably similar to other attempts to classify states related to learning with consumer-grade EEG (i.e., Rafidi et al., forthcoming), there are very few studies to compare the current results to. However, given the wide range of the data between subjects, it can be suggested that further research need focus on flexible implementations relevant to signal classification *per individual*. Chiefly, more advanced feature extraction methods, such as independent components analysis, need be implemented properly for the purposes of future exploration. Moreover, future work in the field of consumer EEG neuofeedback should focus on developing procedure and analytical techniques which reduce the need for between-subject comparisons to be made.

Future directions of the current paradigm include most prominently the implementation of a classifier which can go beyond the binary classification procedures discussed in these results and towards a weighted, 4-class classifier which would provide much more rich data with which to potentially modulate neurofeedback levels. Multi-class classification in neurofeedback represents a mostly yet-to-be-explored domain of the field and promises more robust and effective neurofeedback intervention techniques.

Power bands most relevant to classification in this study appear to align with past investigations. Classification performance measures demonstrated in the current project tentatively support the role of oscillations in memory formation (e.g., Scholz, Schneider, & Rose, 2017). Specifically, these data support the role of theta frequencies in memory formation (see Figure R12; Fell et al., 2011; Staudigl & Hanslmayr, 2013; Sederberg et al., 2013). Klimesch, Doppelmayr, Pachinger, & Ripper (1997) also found similar results for implicit subsequent memory effects, who suggest the possible relationship of this theta-band activity and hippocampal theta induced in the cortex via hippocampo-cortical feedback loops. These data also seem to align with attempts to classify single-trial subsequent memory effects which evidence a robust involvement of high beta activity (e.g., Noh, Herzmann, Curran, & de Sa, 2014). While no association between memory states and gamma activity were evidenced in the current results (as in Sederberg et al., 2003), this may be due to the notch filter cutting out high gamma frequencies, suggesting an explicit disadvantage of the materials utilized in this particular study.

Likewise, the alpha and theta band associated with successful attention classification have been previously evidence in the literature, particularly for sustained attention like in the current study (see Figure R13; e.g., Behzadnia, Ghoshuni, & Charmahini, 2017).

### *Limitations*

The amSMART was developed for the purpose of determining whether the correlates of attention status and memory status as attained with consumer-grade EEG hardware are independently meaningful for classification. While the amSMART appears to have successfully manipulated attention and categorized memories in a tentatively EEG-detectable way, this paradigm has several notable limitations that should be addressed in future work.

While attention may have been manipulated and attentional correlates recorded, the question yet remains whether this kind of attentional process is generalizable to other forms of learning. Additionally, attention may have waned across the eight second trials, as evidenced by the general decrease in classifier accuracy as the trial epochs progressed (see Figure R11). As has been discussed in this paper thus far, attention is not a unitary process, nor is it inextricable from the types of neurological manipulation attentional networks employ. Thus, to some extent, these correlates reflect processes specific to the literal task as hand, potentially down to the color that is being selected for. Future work need explore and create a variety of systems for recording attentional processes that are as relevant to those in the learning environments intended for as possible. While visual attention is certainly important if one wishes to read a textbook better, or faster, but if one wishes to better absorb, say, a recorded lecture, neurofeedback based solely on visual attention engagement would likely be unsuccessful in achieving effective results.

Likewise, memory correlates were attained via a combination of recognition processes. While this may be useful as a preliminary stage, it is doubtful that the correlates of image recognition processes are identical to those used in, for example, recalling detailed information about conceptual material—as one might try to do if this system were to be implemented in naturalistic learning environment such as during studying—or for rote memorization. While in the main, these data may reflect cognitive processes specific to memory, the question remains whether the kind of memory being utilized in the current paradigm is useful for more complex forms of learning for which the system was initially intended. Training a BCI for learning neurofeedback based on data that represents simple image recognition processes may be helpful if, say, the test that's being studied for is an image recognition test; but to some extent, these data must reflect basic perceptual processes at play in recognizing an image that are not useful for conceptual learning processes. For example, having a category lure present during Phase 2 of the amSMART (recognition test) may actually decrease the effectiveness of neurofeedback for complex learning because what's being tested for in such a scenario are the differences in perceptual features between the correct image and lure; it's not as likely trained on conceptual features due to the nature of the test, which rewards perceptual discrimination and punishes scene image memories encoded conceptually rather than perceptually (e.g., if a user remembers that they had been seen a bike, but nothing else, a bike lure would be selected 50% of the time, while if a user remembers the lines and shapes and shadows of the image—for example—but not the actual content of the image, this would be rewarded). For much learning, conceptual features are of priority, thus future investigations need carefully construct paradigms most relevant to the types of content one attempts to remember when learning in daily life.

Additional limitations of this paradigm include the necessity for regular motor responses during the recording of EEGs, which may have contaminated results and reduce classifier performance. Furthermore, participants in the current study completed differential numbers of visual search trials across condition (average visual search arrays responded to during a difficult visual search series = 4.73; average visual search arrays responded to during an easy visual search series = 8.12). This may have impacted classifier performance in many ways because on the one-hand, these motor results are contaminated the EEGs, but they are also predictive of the attention condition. Future work should examine the role of these motor artifacts in EEG classification performance.

This paradigm was intended as a first step in a larger project, as has been discussed. And a realization of this broader goal would mitigate many of these limitations. The amSMART was intended to essentially act as training wheels for a closed-loop neurofeedback paradigm which automatically updates its algorithmic predictions of learning states as the user continues to engage with it (see Figure 22). Thus the training data recorded during the amSMART might be eventually overwritten or down-weighted as a user engages longer with the full neurofeedback paradigm.

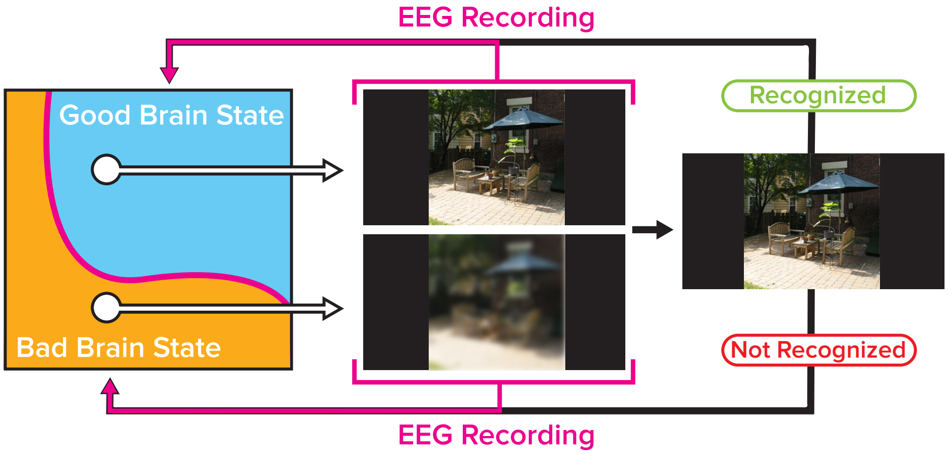


**Figure 22.** A future goal of the amSMART paradigm. Brain states, classified in real-time (left), manipulate the extent to which the image presented to a user is decipherable (center)—rewarding brain states conducive to learning. Recognition tests at the end of each block (right) feed relevant brain data epochs back into the algorithm—weighting high-performance state correlates more highly and down-weighting poor-performance states—improving the classifier's decision boundary and creating an unsupervised, self-updating, closed-loop neurofeedback paradigm for learning.

While preliminary, this study (1) helps to exhibit and clarify some of the difficulties inherent in EEG-based paradigms, (2) presents a novel computer task used for acquiring correlates of attentional and memory processes, (3) provides preliminary data for designing real-time EEG classification of the states produced via the novel task, and (4) exhibits that memory and attentional data recorded with consumer-grade EEG may be classifiable—as would be useful in designing a portable neurofeedback paradigm to improve learning.

## References

Adam, J. J., Davelaar, E. J., Van der Gouw, A., & Willems, P. (2008). Evidence for attentional processing in spatial localization. *Psychological Research*, 72(4), 433-442.

Addante, R. J., Watrous, A. J., Yonelinas, A. P., Ekstrom, A. D., & Ranganath, C. (2011). Prestimulus theta activity predicts correct source memory retrieval. *Proceedings of the National Academy of Sciences*, *108*(26), 10702–70707.

Altmann, E. M., & Gray, W. D. (2002). Forgetting to remember: The functional relationship of decay and interference. *Psychological science*, *13*(1), 27-33.

Anderson, B. A., Jacoby, L. L., Thomas, R. C., & Balota, D. A. (2011). The effects of age and divided attention on spontaneous recognition. *Memory & Cognition*, *39*(4), 725-735.

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, *410*, 131–134.

Anderson, N. D., Iidaka, T., Cabeza, R., Kapur, S., McIntosh, A. R., & Craik, F. I. (2000). The effects of divided attention on encoding-and retrieval-related brain activity: A PET study of younger and older adults. *Journal of cognitive neuroscience*, *12*(5), 775-792.

Andersen, P., Morris, R., Amaral, D., Bliss, T., & O' Keefe, J. (2009). *The Hippocampus Book*. New York: Oxford University Press

Annese, J., Schenker-Ahmed, N. M., Bartsch, H., Maechler, P., Sheh, C., Thomas, N., ... & Klaming, R. (2014). Postmortem examination of patient HM's brain based on histological sectioning and digital 3D reconstruction. *Nature communications*, *5*, 3122.

Arns, M., Drinkenburg, W., & Kenemans, J. L. (2012). The effects of QEEG-informed neurofeedback in ADHD: an open-label pilot study. *Applied Psychophysiology and Biofeedback*, *37*(3), 171-180.

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, 16(8), 437-443.

Badcock, N. A., Mousikou, P., Mahajan, Y., De Lissa, P., Thie, J., & McArthur, G. (2013). Validation of the Emotiv EPOC EEG gaming system for measuring research quality auditory ERPs. *PeerJ*, *1*, e38.

Badcock, N. A., Preece, K. A., de Wit, B., Glenn, K., Fieder, N., Thie, J., & McArthur, G. (2015). Validation of the Emotiv EPOC EEG system for research quality auditory event-related potentials in children. *PeerJ*, *3*, e907.

Barkley, R. A., Murphy, K. R., O'Connell, T., & Connor, D. F. (2005). Effects of two doses of methylphenidate on simulator driving performance in adults with attention deficit hyperactivity disorder. *Journal of safety research*, *36*(2), 121-131.

Behzadnia, A., Ghoshuni, M., & Chermahini, S. A. (2017). EEG Activities and the Sustained Attention Performance. *Neurophysiology*, *49*(3), 226-233.

Benjamin, L. T. J. (2000). The psychology laboratory at the turn of the 20th century. *American Psychologist*, *55*(3), 318–321.

Bennett, M., Dennett, M., Hacker, P., & Searle, J. (2009). *Neuroscience and philosophy: Brain, mind, and language*. Columbia University Press.

Benoni, H., & Tsal, Y. (2013). Conceptual and methodological concerns in the theory of perceptual load. *Frontiers in Psychology*, *4*, 522.

Berker, E. A., Berker, A. H., & Smith, A. (1986). Translation of Broca’s 1865 Report: Localization of Speech in the Third Left Frontal Convolution. *Archives of Neurology*, *43*(10), 1065–1072.

Berman, M. H., & Frederick, J. A. (2009). Efficacy of neurofeedback for executive and memory function in dementia. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *5*(4), e8.

Bernecker, S. (2010). Memory: A Philosophical Study. New York: Oxford University Press.

Bernstein, L. J., Beig, S., Siegenthaler, A. L., & Grady, C. L. (2002). The effect of encoding strategy on the neural correlates of memory for faces. *Neuropsychologia*, *40*(1), 86-98.

Bird, J. E. (1976). Effects of intentional and incidental instructions on picture recognition. *Perceptual and Motor Skills*, *42*(2), 555-561.

Blankertz, B., Dornhege, G., Krauledat, M., Müller, K. R., & Curio, G. (2007). The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects. *NeuroImage*, *37*(2), 539–550.

Bravo, M. J., & Nakayama, K. (1992). The role of attention in different visual-search tasks. *Perception & Psychophysics*, *51*(5), 465–472.

Broadbent, D. E. (1958). Perception and communication. London: Pergamon Press

Broca, P. (2003). Loss of speech, chronic softening and partial destruction of the anterior left lobe of the brain. (C. D. Greene Trans.). *Bulletin de la Société Anthropologique*, 2, 235-238. (Original work published 1861)

Buzsáki, G. (2006). Rhythms of the Brain. New York, NY: Oxford University Press.

Cabeza, R., Mazuz, Y. S., Stokes, J., Kragel, J. E., Woldorff, M. G., Ciaramelli, E., ... & Moscovitch, M. (2011). Overlapping parietal activity in memory and perception: evidence for the attention to memory model. *Journal of cognitive neuroscience*, *23*(11), 3209-3217.

Castelhano, M., & Henderson, J. (2005). Incidental visual memory for objects in scenes. *Visual Cognition*, *12*(6), 1017-1040.

Chica, A. B., Martín-Arévalo, E., Botta, F., & Lupiánez, J. (2014). The Spatial Orienting paradigm: How to design and interpret spatial attention experiments. *Neuroscience & Biobehavioral Reviews*, 40, 35-51.

Christopher deCharms, R. (2008). Applications of real-time fMRI. *Nature Reviews Neuroscience*, *9*(9), 720–729.

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*.

Clayton, M. S., Yeung, N., & Cohen Kadosh, R. (2015). The roles of cortical oscillations in sustained attention. *Trends in Cognitive Sciences*.

Clemens, B., Regenbogen, C., Koch, K., Backes, V., Romanczuk-Seiferth, N., Pauly, K., ... & Kellermann, T. (2015). Incidental memory encoding assessed with signal detection theory and functional magnetic resonance imaging (fMRI). *Frontiers in Behavioral Neuroscience*, *9*, 305.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215.

Corkin, S. (1968). Acquisition of motor skill after bilateral medial temporal-lobe excision. *Neuropsychologia*, *6*(3), 255–265.

Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, 125(2), 159.

Dale, B. (2014, January 07). Brain-wave-reading technology creates amazing, thought-provoking art. Retrieved April 2, 2018, from Slate Magazine

DeBettencourt, M. T., Cohen, J. D., Lee, R. F., Norman, K. A., & Turk-Browne, N. B. (2015). Closed-loop training of attention with real-time brain imaging. *Nature Neuroscience*, *18*(3), 470–478.

DeSchepper, B., & Treisman, A. (1996). Visual memory for novel shapes: implicit coding without attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 27.

Deutch, J. A., & Deutch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, 70, 80-90.

Duvinage, M., Castermans, T., Petieau, M., Hoellinger, T., Cheron, G., & Dutoit, T. (2013). Performance of the Emotiv Epoc headset for P300-based applications. *Biomedical engineering online*, *12*(1), 56.

Düzel, E., Penny, W. D., & Burgess, N. (2010). Brain oscillations and memory. *Current Opinion in Neurobiology*.

Ebbinghaus, H. (1913). Memory: A contribution to experimental psychology. *Classics in the History of Psychology,* (H. A. Ruger & C. E. Bussenius, Trans.). (Original work published 1885). Retrieved from http://nwkpsych.rutgers.edu/~jose/courses/578_mem_learn/2012/readings/Ebbinghaus_1885.pdf

Eichenbaum, H. (2011). The cognitive neuroscience of memory: An introduction. Oxford University Press.

Eichenbaum, H. (2013). What H.M. taught us. *Journal of Cognitive Neuroscience*, *25*(1), 14-21.

Ekandem, J. I., Davis, T. A., Alvarez, I., James, M. T., & Gilbert, J. E. (2012). Evaluating the ergonomics of BCI devices for research and experimentation. *Ergonomics*, *55*(5), 592-598.

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, *113*(28), 7900–7905.

Fell, J., Klaver, P., Elfadil, H., Schaller, C., Elger, C. E., & Fernández, G. (2003). Rhinal-hippocampal theta coherence during declarative memory formation: Interaction with gamma synchronization? *European Journal of Neuroscience*, *17*(5), 1082–1088.

Fell, J., Ludowig, E., Staresina, B. P., Wagner, T., Kranz, T., Elger, C. E., & Axmacher, N. (2011). Medial temporal theta/alpha power enhancement precedes successful memory encoding: evidence based on intracranial EEG. *Journal of Neuroscience*, *31*(14), 5392-5397.

Fielenbach, S. M., Donkers, F. C., Spreen, M., & Bogaerts, S. (2017). Neurofeedback training for psychiatric disorders associated with criminal offending: A systematic review. *Frontiers in Psychiatry*, *8*, 313.

Fok, S., Schwartz, R., Wronkiewicz, M., Holmes, C., Zhang, J., Somers, T., … Leuthardt, E. (2011). An EEG-based brain computer interface for rehabilitation and restoration of hand control following stroke using ipsilateral cortical physiology. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 6277–6280.

Fukuda, K., & Woodman, G. F. (2015). Predicting and improving recognition memory using multiple electrophysiological signals in real time. *Psychological science*, *26*(7), 1026-1037.

Goense, J., Bohraus, Y., & Logothetis, N. K. (2016). fMRI at High Spatial Resolution: Implications for BOLD-Models. *Frontiers in Computational Neuroscience*, *10*.

Guderian, S., Schott, B. H., Richardson-Klavehn, A., & Duzel, E. (2009). Medial temporal theta state before an event predicts episodic encoding success in humans. *Proceedings of the National Academy of Sciences*, *106*(13), 5365–5370.

Guler, I., & Ubeyli, E. D. (2007). Multiclass support vector machines for EEG-signals classification. *IEEE Transactions on Information Technology in Biomedicine*, *11*(2), 117-126.

Haber, S. N. (2011). Neuroanatomy of reward: A view from the ventral striatum. *Neurobiology of Sensation and Reward*, 235.

Hanslmayr, S., Staudigl, T., & Fellner, M.-C. (2012). Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. *Frontiers in Human Neuroscience*, *6*.

Headley, D. B., & Paré, D. (2017). Common oscillatory mechanisms across multiple memory systems. *Npj Science of Learning*, *2*(1), 1.

Heffner, H. E. (1987). Ferrier and the Study of Auditory Cortex. *Archives of Neurology*, *44*(2), 218–221.

Heinke, D., & Humphreys, G. W. (2005). Computational models of visual selective attention: A review. *Connectionist Models in Cognitive Psychology*, 273–312.

Hochberg, L. R., Serruya, M. D., Friehs, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., … Donoghue, J. P. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, *442*(7099), 164–171.

Holmes, G. L., & Khazipov, R. (2007). Basic neurophysiology and the cortical basis of EEG. In *The Clinical Neurophysiology Primer* (pp. 19–33).

Horikawa, T., Tamaki, M., Miyawaki, Y., & Kamitani, Y. (2013). Neural decoding of visual imagery during sleep. *Science*, *340*(6132), 639-642.

Hulbert, J. C., Hirschstein, Z., Brontë, C. A., & Broughton, E. (2018). Unintended side effects of a spotless mind: Theory and practice. *Memory*, *26*(3), 306-320.

Hupbach, A., Hardt, O., Gomez, R., & Nadel, L. (2008). The dynamics of memory: Context-dependent updating. *Learning & Memory*, *15*(8), 574-579.

Jerome, L., Segal, A., & Habinski, L. (2006). What we know about ADHD and driving risk: A literature review, meta-analysis and critique. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *15*(3), 105.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. *Attention And Performance*, *9*, 187-203.

Kimchi, R., Yeshurun, Y., Spehar, B., & Pirkner, Y. (2016). Perceptual organization, visual attention, and objecthood. *Vision Research*, *126*, 34–51.

Klimesch, W., Doppelmayr, M., Pachinger, T., & Ripper, B. (1997). Brain oscillations and human memory: EEG correlates in the upper alpha and theta band. *Neuroscience letters*, *238*(1-2), 9-12.

Klink, P. C., Jentgens, P., & Lorteije, J. A. (2014). Priority maps explain the roles of value, attention, and salience in goal-oriented behavior. *Journal of Neuroscience*, *34*(42), 13867-13869.

Ko, L.-W., Komarov, O., Hairston, W. D., Jung, T.-P., & Lin, C.-T. (2017). Sustained attention in real classroom settings: An EEG study. *Frontiers in Human Neuroscience*, *11*, 388.

Krusienski, D. J., Grosse-Wentrup, M., Galán, F., Coyle, D., Miller, K. J., Forney, E., & Anderson, C. W. (2011). Critical issues in state-of-the-art brain–computer interface signal processing. *Journal of neural engineering*, *8*(2), 025002.

Lacot, E., Vautier, S., Kőhler, S., Pariente, J., Martin, C. B., Puel, M., ... & Barbeau, E. J. (2017). Familiarity and recollection vs representational models of medial temporal lobe structures: A single-case study. *Neuropsychologia*, *104*, 76-91.

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & psychophysics*, *56*(2), 183-197.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 451–468.

Lee, J. H., Ryu, J., Jolesz, F. A., Cho, Z. H., & Yoo, S. S. (2009). Brain-machine interface via real-time fMRI: Preliminary study on thought-controlled robotic arm. *Neuroscience Letters*, *450*(1), 1–6.

Linden, D., & Lancaster, T. (2011). Functional Magnetic Resonance Imaging (FMRI)-based neurofeedback as a new treatment tool for depression. *European Psychiatry*, *26*, 937.

Logothetis, N. K. (2002). The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *357*(1424), 1003–1037.

Loo, S. K., & Barkley, R. A. (2005). Clinical utility of EEG in attention deficit hyperactivity disorder. *Applied Neuropsychology*, *12*(2), 64–76.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, *4*(2), R1–R13.

Maskeliunas, R., Damasevicius, R., Martisius, I., & Vasiljevas, M. (2016). Consumer-grade EEG devices: are they usable for control tasks?. *PeerJ*, *4*, e1746.

Matthews, P. M., Honey, G. D., & Bullmore, E. T. (2006). Applications of fMRI in translational medicine and clinical practice. *Nature Reviews Neuroscience*, 7(9), 732-744.

McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*, *107*(1), 343–352.

Mccormick, D. A., Lavond, D. G., Clark, G. A., Kettner, R. E., Rising, C. E., & Thompson, R. F. (1981). The engram found? Role of the cerebellum in classical conditioning of nictitating membrane and eyelid responses. *Bulletin of the Psychonomic Society*, *18*(3), 103-105.

McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological review*, *39*(4), 352.

Min, J., Wang, P., & Hu, J. (2017). Driver fatigue detection through multiple entropy fusion analysis in an EEG-based system. *PLoS one*, *12*(12), e0188756.

Mitchell, T. M., Hutchinson, R., Niculescu, R. S., & Wang, X. (2004). Learning to decode cognitive states from brain images. *Machine Learning*, *57*, 145–175.

Modha, D. S. (n.d.). Introducing a Brain-inspired Computer. Retrieved April 8, 2018, from http://www.research.ibm.com/articles/brain-chip.shtml

Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., ... & Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. *Journal of anatomy*, *207*(1), 35-66.

Murphy, G., Groeger, J. A., & Greene, C. M. (2016). Twenty years of load theory—Where are we now, and where should we go next? *Psychonomic Bulletin and Review*, *23*(5), 1316-1340.

Naveh-Benjamin, M., Guez, J., Hara, Y., Brubaker, M. S., & Lowenschuss-Erlich, I. (2014). The effects of divided attention on encoding processes under incidental and intentional learning instructions: Underlying mechanisms?. *The Quarterly Journal of Experimental Psychology*, *67*(9), 1682-1696.

Noh, E., Herzmann, G., Curran, T., & de Sa, V. R. (2014). Using single-trial EEG to predict and analyze subsequent memory. *NeuroImage*, *84*, 712-723.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, 10, 424 – 430.

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, *87*(24), 9868-9872.

Otten, L. J., Quayle, A. H., Akram, S., Ditewig, T. A., & Rugg, M. D. (2006). Brain activity before an event predicts later recollection. *Nature Neuroscience*, *9*(4), 489–491.

Paller, K. A., Kutas, M., & Mayes, A. R. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalography and Clinical Neurophysiology*, *67*(4), 360–371.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3-25.

Ptak, R. (2012). The frontoparietal attention network of the human brain: action, saliency, and a priority map of the environment. *The Neuroscientist, 18*(5), 502-515.

Quirk, D. A. (1995). Composite biofeedback conditioning and dangerous offenders: III. *Journal of Neurotherapy*, *1*(2), 44-54.

Radvansky, G. A. (2011). *Human memory* (2nd ed.). London: Routledge.

Rafidi, N. (2012). A brain-computer interface for enhanced learning using classification of EEG data. *Princeton University*.

Rafidi, N., Hulbert, J. C., Pacheco, P., & Norman, K. (forthcoming). Reductions in retrieval competition predict the benefit of repeated testing.

Razavi, B., O'Neill, W. E., & Paige, G. D. (2007). Auditory spatial perception dynamically realigns with changing eye position. Journal of Neuroscience, 27(38), 10249-10258.

Read, S. J., Cesa, I. L., Jones, D. K., & Collins, N. L. (1990). When Is the federal budget like a baby? Metaphor in political rhetoric. *Metaphor and Symbolic Activity*, *5*(3), 125–149.

Reddy, R. P., Rajan, J., Bagavathula, I., & Kandavel, T. (2009). Neurofeedback training to enhance learning and memory in patient with traumatic brain injury: A single case study. *International Journal of Psychosocial Rehabilitation, 14*, 21–28.

Reeves. (2008). Disorders of the nervous system. Retrieved April 2, 2018, from https://www.dartmouth.edu/~dons/figures/chapt_2/Fig_2_1.htm

Remington, R. W., & Folk, C. L. (2001). A dissociation between attention and selection. *Psychological Science*, *12*(6), 511-515.

Ribot, T. (1892). Memory. In D. H. Tuke (Ed.), A dictionary of psychological medicine. Philadelphia: Blakiston.

Romer D., Lee, Y.-C., Mcdonald, C. C.,Winston, F. K. (2014). Adolescence, attention allocation, and driving safety. *The Journal of Adolescent Health.*

Ros, T., Moseley, M. J., Bloom, P. A., Benjamin, L., Parkinson, L. A., & Gruzelier, J. H. (2009). Optimizing microsurgical skills with EEG neurofeedback. *BMC Neuroscience*, *10*(1), 87.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in cognitive sciences*, *11*(6), 251-257.

Rugg, M. D., Fletcher, P. C., Frith, C. D., Frackowiak, R. S., & Dolan, R. J. (1997). Brain regions supporting intentional and incidental memory: a PET study. *Neuroreport*, *8*(5), 1283-1287.

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends in cognitive sciences*, *4*(3), 108-115.

Salari, N., & Rose, M. (2016). Dissociation of the functional relevance of different pre-stimulus oscillatory activity for memory formation. *NeuroImage*, *125*, 1013–1021.

Sanquist, T. F., Rohrbaugh, J. W., Syndulko, K., & Lindsley, D. B. (1980). Electrocortical Signs of Levels of Processing: Perceptual Analysis and Recognition Memory. *Psychophysiology*, *17*(6), 568–576.

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: remembering, imagining, and the brain. *Neuron*, *76*(4), 677-694.

Scholz, S., Schneider, S. L., & Rose, M. (2017). Differential effects of ongoing EEG beta and theta power on memory formation. *PloS one*, *12*(2), e0171913.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery & Psychiatry*, *20*(1), 11–21.

Sederberg, P. B., Kahana, M. J., Howard, M. W., Donner, E. J., & Madsen, J. R. (2003). Theta and gamma oscillations during encoding predict subsequent recall. *Journal of Neuroscience*, *23*(34), 10809-10814.

Sherry, D. F., & Schacter, D. L. (1987). The evolution of multiple memory systems. *Psychological Review*, *94*(4), 439–454.

Shih, J. J., Krusienski, D. J., & Wolpaw, J. R. (2012). Brain-computer interfaces in medicine. *Mayo Clinic Proceedings*, *87(3)*, 268-279.

Shuell, T. J. (1986). Cognitive conceptions of learning. *Review of educational research*, *56*(4), 411-436.

Silva, M. A., See, A. P., Essayed, W. I., Golby, A. J., & Tie, Y. (2017). Challenges and techniques for presurgical brain mapping with functional MRI. *NeuroImage: Clinical.*

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059-1074.

Slotnick, S. D. (2013). The nature of recollection in behavior and the brain. *Neuroreport*, *24*(12), 663-670.

Slotnick, S. D., & Dodson, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory & cognition, 33*(1), 151-170.

Smith, D. V., Davis, B., Niu, K., Healy, E. W., Bonilha, L., Fridriksson, J., ... & Rorden, C. (2010). Spatial attention evokes similar activation patterns for visual and auditory stimuli. *Journal of Cognitive Neuroscience*, 22(2), 347-361.

Sokoloff, L. (2008). The physiological and biochemical bases of functional brain imaging. *Cognitive neurodynamics*, *2*(1), 1-5.

Squire, L. R. (2009). The legacy of patient H.M. for neuroscience. *Neuron*, 61(1), 6-9.

Squire, L. R., & Wixted, J. T. (2011). The cognitive neuroscience of human memory since H.M. *Annual Review of Neuroscience*, *34*(1), 259–288.

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*(5026), 1380–1386.

Staudigl, T., & Hanslmayr, S. (2013). Theta oscillations at encoding mediate the context-dependent nature of human episodic memory. *Current Biology*, *23*(12), 1101-1106.

Sutton, S., Tueting, P., Zubin, J., & John, E. R. (1967). Information delivery and the sensory evoked potential. *Science*, *155*(3768), 1436-1439.

Tank, S. (2013). A brain-computer interface for enhanced learning using classification of EEG data. *Princeton University*.

Tatum, W., Husain, A., Benbadis, S., & Kaplan, P. (2008). *Handbook of EEG Interpretation*. *Medicine*.

Taylor, P. C., Nobre, A. C., & Rushworth, M. F. (2006). FEF TMS affects visual cortical activity. *Cerebral Cortex*, *17*(2), 391-399.

Thompson-Schill, S. L., Swick, D., Farah, M. J., D'Esposito, M., Kan, I. P., & Knight, R. T. (1998). Verb generation in patients with focal frontal lesions: A neuropsychological test of neuroimaging findings. *Proceedings of the National Academy of Sciences*, *95*(26), 15855-15860.

Teplan, M. (2002). Fundamentals of EEG measurement. *Measurement Science Review*, *2*(2), 1–11.

Thompson, W. F., Hall, M. D., & Pressing, J. (2001). Illusory conjunctions of pitch and duration in unfamiliar tone sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 128.

Treisman, A. (1969). Strategies and models of selective attention. Psychological Review, 76, 282-299.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 194–214.

Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*(1373), 1295–1306.

Treisman, A., & Geffen, G. (1967). Selective attention: Perception and response? *Quarterly Journal of Experimental Psychology,* 19, 1-18.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97-136.

Tyler, H. R. (2014). Gall, Franz Joseph. *Encyclopedia of the Neurological Sciences*, 394–394.

Uttal, W. R. (2001). The new phrenology: The limits of localizing cognitive processes in the brain. *Brain and Mind (3)*, 221-228.

Valins, S. (1966). Cognitive effects of false heart-rate feedback. *Journal of Personality and Social Psychology*, *4*(4), 400–408.

van der Veen, F. M., Nijhuis, F. A., Tisserand, D. J., Backes, W. H., & Jolles, J. (2006). Effects of aging on recognition of intentionally and incidentally stored words: An fMRI study. *Neuropsychologia*, *44*(12), 2477-2486.

Vogelsang, D. A., Gruber, M., Bergstrom, Z. M., Ranganath, C., & Simons, J. S. (2017). Alpha oscillations during incidental encoding predict subsequent memory for new "foil" information.

Weeks, J. C., & Hasher, L. (2017). Divided attention reduces resistance to distraction at encoding but not retrieval. *Psychonomic bulletin & review*, *24*(4), 1268-1273.

Weiskopf, N., Scharnowski, F., Veit, R., Goebel, R., Birbaumer, N., & Mathiak, K. (2004). Self-regulation of local brain activity using real-time functional magnetic resonance imaging (fMRI). *Journal of Physiology Paris*.

Wen, D., Wei, Z., Zhou, Y., Li, G., Zhang, X., & Han, W. (2018). Deep learning methods to process fMRI data and their application in the diagnosis of cognitive impairment: A brief overview and our opinion. *Frontiers in Neuroinformatics*, *12*, 23.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 1-8.

Yan, S., Ding, G., Li, H., Sun, N., Wu, Y., Guan, Z., ... & Huang, T. (2016). Enhancing audience engagement in performing arts through an adaptive virtual environment with a brain-computer interface. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, ACM, 306-316.

Yantis, S., & Johnston, J. C. (1990). On the locus of visual selecion: Evidence from focused attention tasks. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 135–149.

Yoo, J. J., Hinds, O., Ofen, N., Thompson, T. W., Whitfield-Gabrieli, S., Triantafyllou, C., & Gabrieli, J. D. (2012). When the brain is prepared to learn: enhancing human learning using real-time fMRI. *Neuroimage*, *59*(1), 846-852.

Zola-Morgan, S. (1995). Localization of Brain Function: The Legacy of Franz Joseph Gall (1758-1828). *Annual Review of Neuroscience,18*(1), 359-383.

Zubin, J., & Barrera, S.E. (1941). Effect of electric convulsive therapy on memory. *Proceedings of the Society of Experimental Biologv and Medicine*, 48, 596-597.

## APPENDICES

# INFORMED CONSENT AGREEMENT

**Protocol number:** 2018FEB07-HIR **Expires:** February 7, 2019
**Study title:** Paying Attention to Real-Time Neurofeedback
**Principal investigator:** Zall Hirschstein

*You are being asked to take part in a research experiment at Bard College that seeks to learn about how different brain states are associated with performance abilities on certain computer-based tasks.*
*To decide whether or not you wish to participate, you should know enough about its risks and benefits to make an informed judgment. This consent form gives you information about the research study, and the experimenter will provide you with additional information about the specific tasks that you will be performing. Once you are ready, you will be asked if you wish to participate and, if so, you will sign the consent form. You can choose not to participate, and you can choose to end your participation at any time during the study.*

**What you will do in this study:** Should you be eligible and decide to participate, you will be asked to make simple judgments about written (words), visual (images), or auditory materials (sounds) presented by a computer by pressing buttons, moving a mouse, or speaking out loud into a microphone that will capture your responses. The researcher will offer detailed instructions to guide you through each part of the experiment and answer any questions you may have about the procedure. After the experiment, you will then be asked to fill in a brief questionnaire about the experiment and given an opportunity to ask any remaining questions that you may have.
During this task, we may record the tiny electrical signals generated by your brain (so-called brainwaves). To do this, small, sterilized electrodes (or ones buffered by clean, disposable felt pads) will be placed over your head using a small amount of gel or saline solution that helps transfer the signal from your body to the recording electrodes, with no risk that they could shock you. The whole process is non-invasive and not painful. You are encouraged keep the researcher informed of your continued comfort during the application of, removal of, and recording using these measurement devices. These data may be used to provide you with feedback about your brain state and may also alter stimuli presented to you on the screen.
It is expected that the first 5-20 minutes of the experiment will be spent preparing you and the measurement devices, leaving the rest of session for the actual task and cleanup. The total time for a session is not expected to run longer than 2 hours. You will be offered the opportunity to take breaks throughout. You may be invited back for additional sessions, but similarly, you can end participation at any time or opt out of future sessions/contacts without penalty. Should you ever decide to end your participation early, you are encouraged to simply let the experimenter know. All the information and responses collected during the experiment will be deleted upon request.

**Risks and benefits:** There are no health risks associated with this study and most participants report having a positive experience. Experiment sessions are kept as short as possible, and every attempt is made to ensure that participants are kept as comfortable as possible throughout. Participants are reminded that, should they become fatigued or in any way uncomfortable during the experiment, they may ask for a break or withdraw at any time without penalty.
After the experiment, participants may prefer, for appearance reasons, to wash off remnants of the completely harmless electrode gel or solution with the provided soap and water.
The words, images, and sounds participants may encounter during the experiment are intended to be neutral, non-threatening, and inoffensive. If you are a student at Bard College and find that any aspect of the experiment caused you distress, you are encouraged to contact the Bard Counseling Center at 845-758-7433 during normal business hours or at 845-758-7777 after hours or on weekends. Even if you are not a Bard College student but find yourself experiencing significant distress, please contact the National

Alliance on Mental Illness (NAMI) at 1-800-950-NAMI (6264).

While this research experiment may not provide participants with any direct benefits, the data collected from this study may help improve the scientific understanding of how to effectively control the focus of attention and the results of doing so. Additionally, we hope that some participants may come away from this experiment with a better grasp of how signals from the brain can influence your everyday life. Moreover, the researchers hope that participants gain insight into the research process at Bard College and beyond through their involvement with this work.

The experimenter will tell you more about the study and our hypotheses at the end of the session.

Compensation: In exchange for participating in this experiment, you may be offered a token piece of candy and raffle entries for Amazon gift cards (with the pot ranging from $25-50), with winners selected at random by May 22, 2017, plus any bonuses introduced during the procedure. Should you be invited back for additional experiment sessions, you will be compensated $5/hour for every hour in these additional sessions.

**Your rights as a participant:** Your participation in this experiment is completely voluntary, and you may withdraw from the experiment at any time without penalty. You will still receive any stated compensation for your participation up until that point. You may withdraw by informing the experimenter that you no longer wish to participate.

**Confidentiality:** All records from this study will be kept confidential. Your responses will be assigned an arbitrary participant number and kept strictly private, shared only with the investigator and trained members of the research team (faculty members and undergraduates at Bard College) who have been certified for work with human participants. We will not include any information that will make it possible to identify you in any report we might publish, including the resulting Senior Project, which will be publicly accessible at Bard College's Stevenson Library and on the online thesis repository, the Digital Commons. Research records will be stored securely in a locked cabinet and/or on password-protected computers.

If you have questions about this study, please ask your researcher, Zall Hirschstein (zh8605@bard.edu), or contact Dr. Justin Hulbert (Psychology Program, Bard College, Annandale-on-Hudson, NY 12504; jhulbert@bard.edu). If you have questions about your rights as a research participant, please contact the Bard College Institutional Review Board at irb@bard.edu.

## STATEMENT OF CONSENT:

***"The purpose of this study, procedures to be followed, and the risks and benefits have been explained to me. I have been given an opportunity to ask questions, and my questions have been answered to my satisfaction. I have been told whom to contact if I have additional questions. I have read this consent form and agree to be in this study, with the understanding that I may withdraw at any time."***

By signing below, I agree with the above **statement of consent** and further certify that I am at least 18 years of age.

Participant signature ______________________ Date ____________

Participant name (printed) ______________________

Experimenter signature ______________________

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 3 | 9 | | | | 5 |
| 9 | | | | | | | 4 | 1 |
| | | 8 | | 4 | 6 | 9 | | 3 |
| 4 | 7 | 6 | | | | | 5 | |
| | | 2 | 6 | | 3 | 4 | | |
| | 8 | | | | | 2 | 1 | 6 |
| 8 | | 4 | 5 | 1 | | 7 | | |
| 2 | 1 | | | | | | | 4 |
| 6 | | | | 3 | 7 | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 8 | 4 | | | | 9 |
| | | 1 | | | | | | 5 |
| 8 | | | | 2 | 1 | 4 | 6 | |
| 7 | | 8 | | | | | 9 | |
| | | | | | | | | |
| | 5 | | | | | 3 | | 1 |
| | 2 | 4 | 9 | 1 | | | | 7 |
| 9 | | | | | | 5 | | |
| 3 | | | | 8 | 4 | | | |

**Study title:** Paying Attention to Real-Time Neurofeedback

**Study Protocol Number:** 2018FEB07-HIR

**Principal investigator:** Zall Hirschstein (zh8605@bard.edu)

Thank you for participating in this experiment. This research is designed to explore the basic mechanisms underlying attentional control and memory. By conducting this study, we hope to learn more about how people might learn to better focus their cognitive state in a way that allows them to remember what they want to remember and when they want to remember.

In the first part of the experiment, we asked you to focus your attention on one or more primary tasks. Although we were interested in how well you performed on the primary task(s) by controlling your attention and the electrophysiological data recorded during this time, we were also interested in your performance and the electrophysiological data associated with memory for stimuli that appeared between, in, or around this attentional task. To examine this, we introduced "distractor" materials and later surprised you with a memory test for these distractors.

The reason for withholding information about the upcoming memory task was that we required a measure of memory for events that took place without the intention for these events to be remembered. By combining the electrophysiological data associated with the attentional manipulation in the first task and the "incidental memory" correlates attained by relating the stimuli you remembered and forgot with the related electrophysiological data, we hoped to present you with neurofeedback that varied between being controlled by your personalized "high attention" brain state correlates and your "high incidental memory" brain state correlates. Because retention of information requires both attentional and memory processes, we hypothesize that ideal neurofeedback for explicit memory would be presented based on some combination of your electrophysiological correlates of these two tasks.

By researching the nature of these combined brain states and utilizing them for feedback, we hope to increase our ability to control our retention for material. For example, students might be able to use this device and computer algorithm studying for an exam.

*This experiment required us to withhold information from you in order to avoid contaminating the results. In particular, we did not tell you in advance about the surprise memory test. Intentionally trying to learn is a very different process than the learning that incidentally occurs when you perform a task. In fact, past research has shown that trying to learn can interfere with more incidental forms of learning. Furthermore, the neurofeedback you observed may not have been related to the purported cognitive processes expressed by your experimenter. This would have been done in order to establish that real neurofeedback has benefits above and beyond that of feedback unrelated to your attentional/memory brain states, and we require some participants to act as a control in order to establish whether our hypotheses are correct. Therefore, telling you up front that you would be tested on these materials and that you may receive sham neurofeedback could invalidate the hypotheses being investigated. We apologize for withholding this information about the experiment before you participated. Please let your researcher know if we may still use your data in our study.*

Regardless, if you have any questions or concerns, you may ask your experimenter, Zall Hirschstein in person or at zh8605@bard.edu, or feel free to contact his faculty supervisor, Dr. Justin C. Hulbert, at jhulbert@bard.edu. You may email the Bard College Institutional Review Board at irb@bard.edu for questions about your rights as a participant.

Again, we thank you for your participation. If you know of any friends or acquaintances that are eligible to participate in this study, we kindly request that you not discuss it with them until after they have had the opportunity to participate. Prior knowledge of questions asked during the study can invalidate the results. We greatly appreciate your cooperation.

| Bard Institutional Review Board | Bard Counseling Center | National Alliance on Mental Illness Hotline |
|---|---|---|
| irb@bard.edu | 845-758-7433 / 7777 | 1-800-950-NAMI (6264) |

**Phase 1:**

- How often did you pay attention to the background “distractor” images while completing the visual search tasks?

| Never | Rarely | Sometimes | Often | Always |
|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 |

- Did you ever pay attention to the background “distractor” images on purpose while completing the visual search tasks?

| Never | Rarely | Sometimes | Often | Always |
|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 |

- To what extent did you expect to be tested for the “distractor” images before/while completing the visual search task?

| Not at all | A little | A bit | Quite a bit | A lot |
|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 |

**Phase 2:**

- Do you think the recognition test captured your memory for the images?

| Not at all | A little | A bit | Quite a bit | A lot |
|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 |

**General:**

- Last night, how many hours of sleep did you get? (estimate)

≤2 3 4 5 6 7 8 9 10 ≥11

- How often did you experience stress completing this experiment?

| Never | Rarely | Sometimes | Often | Always |
|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 |

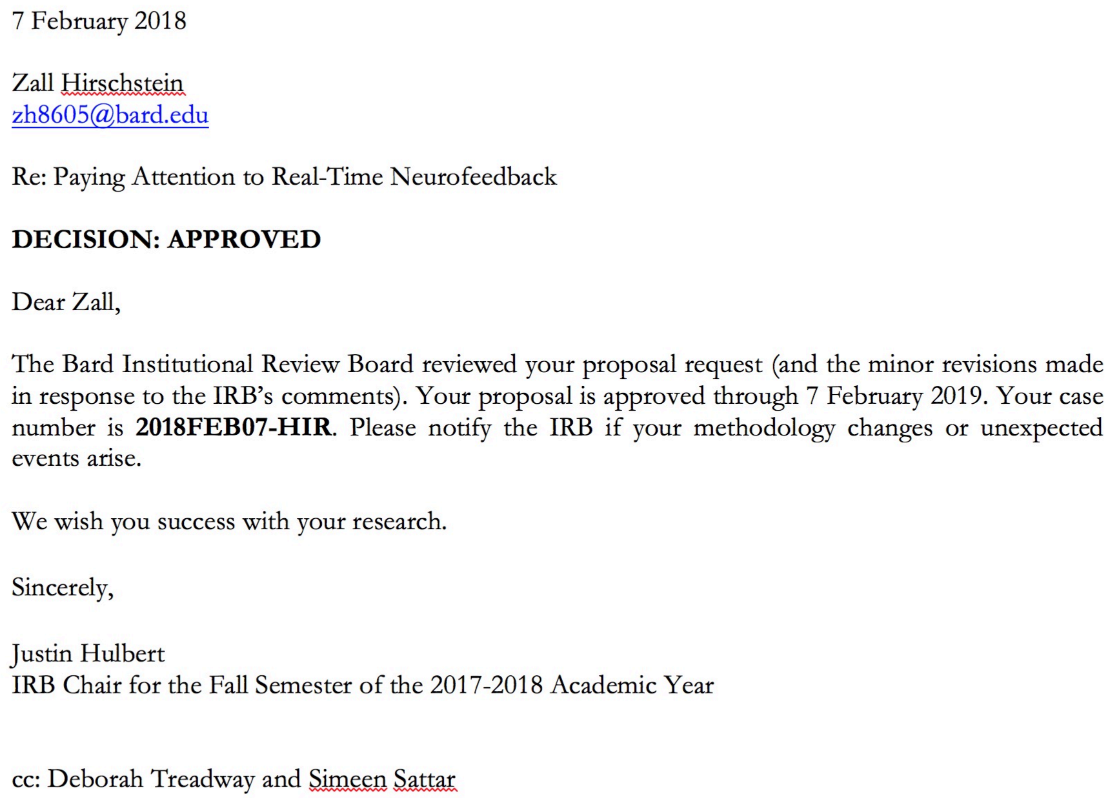
- Do you have other comments or questions?

## Certificate of Completion

The National Institutes of Health (NIH) Office of Extramural Research certifies that **Zall Hirschstein** successfully completed the NIH Web-based training course "Protecting Human Research Participants".

Date of completion: 02/14/2016.

Certification Number: 2003376.



7 February 2018

Zall Hirschstein
zh8605@bard.edu

Re: Paying Attention to Real-Time Neurofeedback

**DECISION: APPROVED**

Dear Zall,

The Bard Institutional Review Board reviewed your proposal request (and the minor revisions made in response to the IRB's comments). Your proposal is approved through 7 February 2019. Your case number is **2018FEB07-HIR**. Please notify the IRB if your methodology changes or unexpected events arise.

We wish you success with your research.

Sincerely,

Justin Hulbert
IRB Chair for the Fall Semester of the 2017-2018 Academic Year

cc: Deborah Treadway and Simeen Sattar