COGNITIVE NEUROSCIENCE EDITOR'S PICK 2021

EDITED BY: Lutz Jäncke
PUBLISHED IN: Frontiers in Human Neuroscience

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88971-174-1
DOI 10.3389/978-2-88971-174-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

COGNITIVE NEUROSCIENCE EDITOR'S PICK 2021

Topic Editor:
Lutz Jäncke, University of Zurich, Switzerland

Citation: Jäncke, L., ed. (2021). Cognitive Neuroscience Editor's Pick 2021. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-174-1

Table of Contents

frontiers in Human Neuroscience

HYPOTHESIS AND THEORY
published: 11 March 2019
doi: 10.3389/fnhum.2019.00028

A Non-cognitive Behavioral Model for Interpreting Functional Neuroimaging Studies

Robert G. Shulman[1,2] and Douglas L. Rothman[1]**

[1] *Magnetic Resonance Research Center, Department of Radiology, Yale University School of Medicine, New Haven, CT, United States,* [2] *Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, United States*

The dominant model for interpreting brain imaging experiments, which we refer to as the Standard Cognitive Model (SCM), assumes that the brain is organized in support of mental processes that control behavior. However, functional neuroimaging experiments of cognitive tasks have not shown clear anatomic segregation between mental processes originally proposed by this model. This failing has been blamed on limitations in imaging technology and non-linearity in the brain's implementation of these processes. However, the validity of the underlying cognitive models used to describe the brain has rarely been questioned or directly tested against imaging results. We propose an alternative model of brain function, that we term the Non-cognitive Behavioral Model (NBM), which correlates observed human behavior directly with measured brain activity without making assumptions about intervening cognitive processes. Our model derives from behavioral psychology but is extended to include brain activity, in addition to behavior, as observables. A further extension is the role of neuroplasticity, as opposed to innate cognitive processes, in developing the brain's support of cognitive behavior. We present the theoretical basis with which the SCM maps cognitive processes onto functional magnetic resonance and positron emission tomography images and compare and contrast with the NBM. We also describe how the NBM can be used experimentally to study how the brain supports behavior. Two applications are presented that support the usefulness of the NBM. In one, the NBM use of the total functional imaging signal (not just the differences between states) provides a stronger correlation of neural activity with the behavioral state of consciousness than the SCM approach in both anesthesia and coma. The second example reviews studies of facial and object recognition that provide evidence for the NBM proposal that neuroplasticity and experience play key roles in the brain's support of recognition and other behaviors. The conclusions regarding neuroplasticity are then generalized to explain the incomplete functional segregation observed in the application of the SCM to neuroimaging.

Keywords: functional magnet resonance imaging, cognitive psychology, neuroenergetics, behavioral psychology, consciousness, object recognition

OPEN ACCESS

Edited by:
Agustin Ibanez, Institute of Cognitive and Translational Neuroscience (INCYT), Argentina

Reviewed by:
Gaël Jobard, Université de Bordeaux, France
Albert Gjedde, University of Southern Denmark, Denmark

****Correspondence:***
Robert G. Shulman
robert.shulman@yale.edu
Douglas L. Rothman
Douglas.rothman@yale.edu

Received: *10 August 2018*
Accepted: *21 January 2019*
Published: *11 March 2019*

Citation:
Shulman RG and Rothman DL (2019) A Non-cognitive Behavioral Model for Interpreting Functional Neuroimaging Studies. Front. Hum. Neurosci. 13:28. doi: 10.3389/fnhum.2019.00028

INTRODUCTION

The ability of imaging to measure reliable, physical chemical properties of neural activity in the brain during a person's behavior created a revolutionary opportunity for psychology. With these new techniques, neural activity was recognized as a novel observable in the study of behavior. While measurements of animal and human behavior previously had been the most reliable observation upon which an understanding of brain function could be built, the reliability of measuring neural activity offered by brain imaging studies is now certainly as trustworthy. Functional magnetic resonance imaging (fMRI), magnetic resonance spectroscopy (MRS), and positron emission tomography (PET) measure the energy consumed by neural activity through glucose and oxygen consumption evaluated by coupled parameters of blood flow and volume. Oxidative glucose metabolism (measured by fMRI and PET) quantitatively tracks both neural electrical activity and glutamate/GABA neurotransmitter release (Rothman et al., 2011; Hyder et al., 2013a,b; Yu et al., 2017).

These and other available measurements of brain functional activity (e.g., EEG, MEG) bear upon a question often asked in psychology as to the role of an individual's internal processes in behavior. The role of internal processes in psychology became prominent 60 years ago when behaviorism, the then popular version of psychology, was severely challenged for having ignored the inner workings of the person by having insisted that only observable behavior could be the source of scientifically valid information. Particularly influential were Noam Chomsky's criticisms of behaviorism for having neglected internal processes during the development of language, which he claimed specifically needed contributions beyond measureable external influences and behavior (Chomsky, 1959). Chomsky (1971) proposed that psychology must include innate, internal mental processes, a position that has dominated the field ever since. Subsequently many others introduced models of cognition based on underlying mental processes. The most extreme version of this theory treats the processes, referred to as mental modules as totally isolatable both in terms of information processing and as implemented in the brain, (Cosmides, 1980; Fodor, 1983). However, although such an extreme degree of independence is rarely present in theories today at least some degree of separability of cognitive processes is almost always assumed (Gazzaniga and Mangun, 2014).

The measurement of direct brain correlates of innate internal mental processes remained elusive until the late 1980s with the introduction of PET functional imaging (Phelps et al., 1979; Fox et al., 1986) and soon after fMRI (Tank et al., 1992). The initial PET studies looked for functional segregation within the brain initially in the tradition of localizationist pioneers such as Broca, Weirnicke, etc. who, based on brain lesions, assigned different brain regions to language and other functions (for a review, see Friston, 2005 and references therein). The formal application of cognitive psychology to functional neuroimaging was introduced by Posner and Raichle (1994) and Posner and Raichle (1998) using an approach based upon the subtractive method introduced in the previous century by Donders (1969) which compared response times as a function of task complexity. The method was soon introduced into the SPM statistical methodology by Friston (1997) and Frackowiak et al. (2004; Ashburner, 2012). Although since that time there have been many different statistical and experimental approaches applied to functional imaging almost all share the assumption that the brain supports pre defined cognitive processes that depend on a degree of functional segregation. We refer to models assuming underlying cognitive processes, derived from cognitive psychology, as the Standard Cognitive Model (SCM).

Despite the continued extraordinary growth of fMRI, which has transcended the initial psychological and neurological applications to move into the social sciences (e.g., neuroeconomics) and popular culture (Satel and Lilienthal, 2013), there has been increasing concern about the disagreements between expectations of cognitive theories and experimental fMRI data (e.g., Friston et al., 1996; Shulman, 1996; McGonigle et al., 2000; Poldrack, 2006; Friston, 2009; Gonzalez-Castillo et al., 2012). In particular clear functional segregation of cognitive processes has been elusive. Most of this criticism has focused on technical issues such as imaging quality and statistical analysis, with additional work focusing on interactions between regions and non-linearity between cognitive processes and the brain's support of them (Price and Friston, 2007; Friston and Price, 2011). In contrast there has been very little criticism of the psychological assumptions embedded in the studies (Shulman, 1996; Shulman and Rothman, 1998; Uttal, 2001; Suhler and Churchland, 2011; Shulman et al., 2014). Shulman (1996) proposed that the use of fMRI to localize brain regions to previously determined mental processes was premature and the opportunity to use functional neuroimaging to develop and test new theories was being neglected. In subsequent papers, we further elaborated this criticism using specific examples from the literature in which the expectations of clear functional segregation of cognitive processes were not being met by a modular brain (Shulman, 1996; Shulman and Rothman, 1998; van Eijsden et al., 2009). However, the goal of using functional neuroimaging to localize or find the patterns of activity that support assumed mental processes if anything has become more dominant. A recent survey of the literature from 2007 through 2011 by Tressoldi et al. (2012) found that only 11% of fMRI studies actually tested cognitive theories, the rest being used only for localization of the assumed processes. Of the 11% only a few met the criteria they set for rigorous testing of models (Tressoldi et al., 2012). Although testing of cognitive theories is not the sole value of neuroimaging studies the relatively few papers with this goal support our contention that assuming the brain has anatomical representations of cognitive concepts has rarely been questioned in imaging studies.

The question of whether the brain supports separable cognitive processes would be moot if fMRI studies showed consistent reproducible localizations that could be assigned to specific cognitive processes independent of context. However, as has been pointed out this expectation has not been met (e.g., Shulman, 1996; van Eijsden et al., 2009). The majority of approaches to address this problem have looked at modifying

the statistical methods by which brain regions associated with cognitive processes are localized (Friston et al., 1996; Price and Friston, 2007) as well as broadening the criteria of functional segregation to accept a considerable amount of anatomical overlap between regions supporting different concepts. We have taken a different approach and instead have proposed that the underlying psychological assumptions regarding the cognitive structure of the brain used in designing and analyzing functional neuroimaging experiments need to be re-examined (Shulman et al., 2009, 2014, Shulman, 2013). To explore alternatives to the top down SCM approach we have proposed that neuroimaging data of brain activities should be directly correlated with behavioral observations without assuming underlying separable mental processes.

The primary goal of this paper is to explicate and formalize this approach, which we refer to as the Non-cognitive Behavioral Model (NBM) to allow its broader use in designing and interpreting functional neuroimaging studies. We start by describing assumptions that underly the application of cognitive theories to functional neuroimaging (see the section "Basic Structure of Standard Cognitive Models (SCM) and Their Application to Functional Neuroimaging"). The lack of agreement of functional neuroimaging data with the expectations of finding functionally segregated support of cognitive processes is described along with the modifications of the SCM assumptions of how the brain supports cognitive concepts to obtain better agreement with experimental data. In Section "The Non-cognitive Behavioral Model (NBM)," we describe the NBM and compare and contrast it with the SCM. Key differences include no assumptions in NBM of underlying cognitive processes, the incomplete separability of brain activity and behavior, and the importance of neuroplasticity and experience in determining patterns of brain activity. Another key difference, specific to neuroimaging, is that the total as opposed to the difference in the functional neuroimaging signal is analyzed. In Section "Application of the NBM to Studies Determining Neural Correlates of Consciousness," we compare the NBM and SCM approaches for localizing neural correlates of consciousness from imaging data. The NBM approach of looking for correlations between the total activity of all brain regions and the measured behavior, provides stronger correlations than between brain regions proposed by cognitive theories to support consciousness and the average cortical activity.

In Section "fMRI Studies of Facial and Object Recognition," recent studies on the fMRI responses to faces in the fusiform gyrus support the important role of experience and neuroplasticity in the development of brain responses. In Section "Application of the NBM to Study Cognitive Behaviors," we generalize the NBM approach and point out the potential key role of neuroplasticity in explaining both similarities and differences between and within individuals performing ostensibly the same behavior. In Section "Epistemological Basis of the NBM," we describe the epistemological basis of the NBM. We conclude by suggesting how NBM provides a useful approach for studying the brain support of cognitive and other behaviors without assuming an underlying cognitive theory.

BASIC STRUCTURE OF STANDARD COGNITIVE MODELS (SCM) AND THEIR APPLICATION TO FUNCTIONAL NEUROIMAGING

The standard methodology used in interpreting functional neuroimaging experiments is based on the assumption, from cognitive psychology, that the brain supports cognitive and other behaviors by integrative processing of separate mental processes. The brain has regions dedicated to supporting these processes, referred to as functional segregation (Frackowiak et al., 2004). The actual behavior that takes place is the result of the functional integration of these separate cognitive processes (Friston, 2005; Price and Friston, 2007), for a comprehensive description of modern cognitive theories, see Gazzaniga and Mangun (2014). For the example of memory, there are different cognitive processes supporting long term, short term, and working memory, and within these grosser processes many sub-processes have been proposed (Baddeley and Hitch, 1974; Goldman-Rakic, 1995; Wager and Smith, 2003; Roth and Courtney, 2007; Jonides et al., 2008; Shulman, 2013, chapter 5; Gazzaniga and Mangun, 2014; Shulman et al., 2014).

The concept of the brain supporting separable cognitive processes has been criticized primarily from a philosophical perspective (e.g., Fodor, 2000; Suhler and Churchland, 2011). However, in neuroimaging it is broadly accepted with the caveat that the brain's implementation of cognitive processes may involve considerable overlap and non-linearity (Friston et al., 1996; Friston and Price, 2011). We briefly describe below a common strategy for using the SCM in functional neuroimaging, primarily in order to emphasize how hypothesized mental processes and sub processes are integral to the experimental design and the interpretation of functional neuroimaging data. We then discuss the limitations that have been found in assigning cognitive processes to unique patterns of brain activity using neuroimaging and the approaches being taken within the SCM paradigm for addressing them. In Section "The Non-cognitive Behavioral Model (NBM)," we describe the NBM as an alternate approach for addressing structure function relationships by dropping cognitive processes as a starting assumption and instead taking a bottom up approach.

Localization of Mental Processes by Functional Neuroimaging

A schematic diagram of how the brain is assumed to be functionally organized in the SCM is shown in **Figure 1**. There are three fundamental components: measured behavior at the top level, mental processes supported by brain activity at an intermediate level and regional brain neuronal activity at the lowest level. The interaction of the environment with the subject (including any sensory stimulation or psychological tasks given to them during the imaging studies) is shown as a lower level input to the brain although the actual situation is more complex including feedback from the subject's own responses as well as any internal behavior. Brain neuronal activity is organized to support

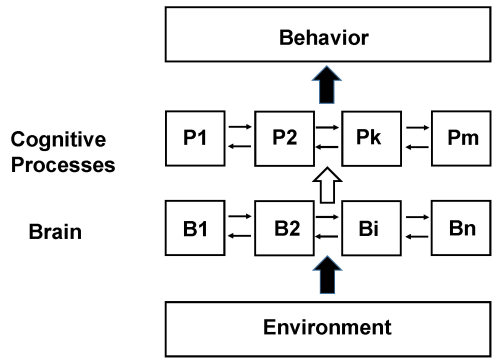


FIGURE 1 | Schematic of the Standard Cognitive Model (SCM). Schematic diagram of the standard model underlying the large majority of functional neuroimaging studies especially as applied to complex behaviors involving cognition and emotion, which we refer to as the SCM. N localized regions of the brain (B1 to Bn) are organized to support m cognitive processes (P1 to Pm). The actual number and function of the theorized cognitive processes as well as the details of their interconnections depend on the specific model used. The cognitive process abstraction allows cognitive behavior to be studied at the level of information processing without reference to how specifically the brain supports it (similar to a computational algorithm not depending upon the computer hardware it is implemented on) (Baddeley and Hitch, 1974; Fodor, 1983; Pinker, 1997). In early functional neuroimaging experiments it was assumed that the brain was organized such that there was sufficient functional segregation to allow a 1 to 1 mapping between discrete brain regions and specific processes (e.g., between B1 and P1). Subsequently due to disagreements found between functional neuroimaging data the assumption of 1:1 mapping this assumption was relaxed allowing linear and non-linear interactions between different brain regions in supporting specific cognitive processes (e.g., between B1, B2, and B7 in their support of P1 and the modulation of their activity if other processes are active) (Friston et al., 1996). Experimental variables are the input to the subject being studied (referred to broadly as Environment but in many studies consisting primarily of an investigator presenting tasks or stimuli to the subject). Measurements are observable behavior and the pattern of brain neuronal activity, usually differential activity in a region relative to some other state, determined from functional neuroimaging.

these mental processes and are assumed to have some degree of functional segregation although considerable interaction between regions may be incorporated (Friston et al., 1996). The functional integration of the outputs of the cognitive processes leads to behavior or perception.

In a typical experiment a subject in an MRI or PET scanner will perform a variety of tasks (e.g., remembering lists of words) and/or be exposed to stimuli designed to differentially activate cognitive processes. The functional neuroimaging signal (in response to an experimental input j) for each voxel is described in analysis packages as a sum of the imaging signal contributions from the region used to support the mental processes engaged to perform the task. However, the relationship between the signal, the neuronal activity underlying it, and how much of the neuronal activity is supporting a cognitive process, is usually left undefined.

In order to clarify these relationships we express the total neuronal activity (N_{ij}) induced in a voxel or region i (see **Figure 1**) by a task j as the sum of the neuronal activity within that voxel supporting each separate mental process Xk (see also the previous description in Shulman et al., 2014).

$$Nij = boXo + \sum_{k=1} bijkXk \quad (1)$$

The value of Xk is nominally set to 1 (see the GLM example below). The actual value of these constants cannot be derived from a bottom up approach so that the relationship between the cognitive process and regional neuronal activity supporting it is entirely empirical. If a cognitive process k is not supported by brain activity in voxel i during task j then bijk = 0. The term boXo refers to neuronal activity assumed to not support any cognitive process. This activity is usually assumed to be equal to the neuronal activity in the voxel when no cognitive task is being performed, and is often referred to as the resting state activity. As described below quantitative imaging studies have found that the size of the term boXo is generally an order of magnitude larger than the incremental activity believed to support cognitive processes.

The relationship between the imaging signal and neuronal activity for fMRI is highly dependent on experimental methodology as well as neurophysiological couplings between blood flow, glucose oxidation, and neuronal signaling during the task (Hoge et al., 1999; Hyder and Rothman, 2012). However, in order to focus on the relationship between neuronal activity and the imaging signal we will assume that the imaging method can be corrected for vascular and metabolic response functions and calibrated such that there is a direct relationship between neuronal activity (e.g., number of spikes per second in an ensemble or number of neurotransmitter quanta released) and the signal measured and equation 1 can be used. These corrections have been performed for fMRI and PET CMRglc measurements of energy consumption[1] (Sibson et al., 1998; Hyder et al., 2013a,b).

Localization of Mental Processes Using Linear Models

In the original PET and fMRI functional neuroimaging studies the series of equations that can be generated from equation 2 were solved through a method sometimes referred to as cognitive subtraction. The primary assumption of cognitive subtraction was based upon the concept in cognitive psychology of pure insertion (Donders, 1969) – that separable mental processes exist and to a first order are not influenced by the activity of other mental processes. Cognitive subtraction was replaced by the use of general linear models (Friston et al., 1996; Friston, 1997) in order to provide a valid statistical framework for

[1]It has been shown by comparison to electrical recording and ^{13}C MRS measurements of glutamate and GABA release and recycling that over a wide interval neuronal activity is proportional to the incremental energy that the brain needs above non-signaling processes (Sibson et al., 1998; Hyder et al., 2013a,b). Therefore methods such as FDG-PET and calibrated fMRI have the potential of providing quantitative maps of neuronal signaling (e.g., Stender et al., 2015, 2016). In order to not obscure the relationships between brain activity and behavior due to issues regarding the veracity with which neuronal signaling is tracked by imaging we assume in this paper (except where stated otherwise) that imaging is providing whole brain maps of neuronal signaling (or the difference in signaling between two states as with fMRI).

assessing the certainty with which the imaging results supported localization of brain activity. Subsequently there has been a move away from linear models which have been criticized (see the section "Explanations for the Lack of Clear Functional Segregation in Cognitive Neuroimaging") primarily on the grounds that it assumes linearity in how the brain supports cognitive processes (Friston et al., 1996; Poldrack, 2006; Price and Friston, 2007). We briefly describe the general linear model below as used in functional neuroimaging below using the mathematical framework introduced above in order to clarify the neuronal basis of both the subsequent modifications within the SCM and how it differs from the NBM description.

General Linear Model (GLM) in Functional Neuroimaging

The concept of cognitive subtraction has largely been performed using statistical packages based upon a mathematical description called the General Linear Model (GLM) which was formally introduced into functional neuroimaging by Friston et al. (1996) and Friston (1997). It remains a major method used for the analysis of fMRI studies although as discussed below many modifications as well as alternate methods are now in use. We briefly describe its application here in order to help illustrate the differences between the underlying assumptions and experimental applications and analyses of the SCM with the NBM described in the Section "The Non-cognitive Behavioral Model (NBM)."

In a standard implementation of a GLM to functional neuroimaging the relationship between the signal in voxel i during task j (Nij) and the underlying cognitive processes (Xk) is assumed to be linear (e.g., the value of bijk in equation 1 is a constant). Furthermore, the brain activity supporting the different cognitive processes do not interact. Therefore the individual terms can be isolated through fitting for the values bijk provided that enough different tasks/stimuli (j) are performed.

As an example of this procedure a simplified working memory model based on the pioneering work of Baddeley and Hitch (1974), consisting of a phonological loop (PL), visual sketch pad (VS), and central processor (CP) is shown in **Figure 2**. If the goal of the experiment is to locate where these processing modules the system is modeled, using a procedure referred to as the design matrix as consisting of three cognitive processes XCP, XPL, XVS. The imaging signal time course is then measured during three (or more) tasks, with the focus being on the time intervals where the contribution from each process is approximately constant. As an example the first imaging measurement is made during an auditory task involving the PL, the second during a task that activates the visual spatial sketch pad but not using the PL and the third during a task that does not activate the visual scratch pad. The signal measured during each task is then modeled with the following series of linear series of equations (equations 2–4) along with an error term ej.

$$\text{Task 1 Ni1} = \text{b0X}_\text{o} + \text{bi11X}_\text{CP} + \text{bi12X}_\text{PL} + \text{bi13X}_\text{VS} + \text{e1} \quad (2)$$

$$\text{Task 2 Ni1} = \text{boX}_\text{o} + \text{bi21X}_\text{CP} + 0 + \text{bi23X}_{\text{VS}+}\text{e2} \quad (3)$$

$$\text{Task 3 Ni1} = \text{boX}_\text{o} + \text{bi31X}_\text{CP} + \text{bi32X}_\text{PL} + 0 + \text{e3} \quad (4)$$

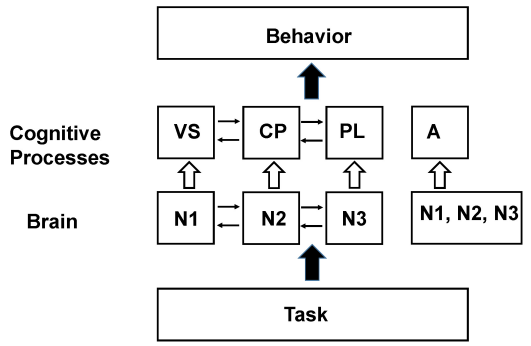


FIGURE 2 | Schematic of the Baddeley and Hitch (1974) model of working memory. In their model working memory consists of a central processor (CP) that performs manipulation of data stored in a visual spatial sketch pad (VS) or a phonological loop (PL). By performing appropriate functional imaging studies the locations in the brain that support these processing modules (N1, N2, N3) can be identified provided that the assumptions in the SCM hold (see the section "Explanations for the Lack of Clear Functional Segregation in Cognitive Neuroimaging"). The resting state neuronal activity is not shown in the diagram since it has been normalized out of the imaging data. However, later functional neuroimaging studies found that these same brain regions are also recruited to support behavioral tasks hypothesized to be unrelated to working memory processes. The violation of the original assumptions about functional segregation of mental processes in the brain has led to the broadening of the assumption to allow a region to support multiple processes (Friston et al., 1996) as well as the introduction of additional processes to better fit behavioral (and sometimes imaging) data such as the concept of attention (A) which was found to activate similar regions in the frontal lobes as working memory tasks (see Shulman, 1996). In an SCM analysis the individual processing modules, N1, N2, or N3 are the experimental parameters to be varied when seeking reproducible correlations between the behavioral components and the brain regions assigned to the same processing module. The difference between these experimental parameters and those that would be used for testing the NBM model are discussed in Section "The Non-cognitive Behavioral Model (NBM)."

Since the neuronal activity supporting a process is assumed to be constant the series of equations above can be solved using standard statistical matrix methods (Friston, 1997) for the constants bijk allowing the regional neuronal support of each process to be localized along with an estimate of the uncertainty in the assignment.

The determination of the coefficients is further simplified by the assumption that only the change measured in the imaging signal (and underlying neuronal activity) during a task or stimulus is relevant for supporting cognitive processes (Morcum and Fletcher, 2007). There are many methods for removing this signal, but the most commonly used equate it to the resting state (or baseline) signal when the subject receives no stimulation or performs tasks (see the section "Explanations for the Lack of Clear Functional Segregation in Cognitive Neuroimaging"), which is equivalent to removing the term boXo in equation 1.

Explanations for the Lack of Clear Functional Segregation in Cognitive Neuroimaging

Since the initial cognitive neuroimaging studies of Posner and Raichle (1994) and Posner and Raichle (1998) there has been

a general expectation that functional segregation of mental processes would be found in the brain. However, there have been a large number of reports in which reproducibility and degree of functional segregation within and between subjects for the same task has not been high (Friston et al., 1996; McGonigle et al., 2000; Otzenberger et al., 2005) particularly for cognitive but also for sensory tasks. Despite this problem studies that find different spatial responses assigned to the same function or similar spatial patterns of activation assigned to the same function have rarely led to questioning of the fundamental cognitive concepts being localized in the brain (for specific examples, see references in Shulman, 1996, 2013; Shulman et al., 2007; van Eijsden et al., 2009). Instead they have been mainly attributed to limitations in the imaging methods and incomplete functional segregation, and linearity, in how the brain implements cognitive processes (Friston et al., 1996; McGonigle et al., 2000; Gonzalez-Castillo et al., 2012). We briefly describe some of the main concerns in this section and how they have been addressed by others in the field. In Section "The Non-cognitive Behavioral Model (NBM)," we describe our alternative proposal that the assumption that the brain supports separable mental processes that can be described in terms of information processing as in cognitive psychology needs to be examined critically.

Influence of Resting State Neuronal Activity on the Functional Neuroimaging

During the initial decade of PET and fMRI functional neuroimaging it was widely believed that despite the brain being energetically very expensive (approximately 20% of the body's oxygen consumption at rest) neuronal signaling activity was not energetically costly. This conclusion was based on several lines of evidence including measurements suggesting low energy costs of action potentials extrapolated to the human brain (Creutzfeldt, 1975), findings from PET of a very small energetic cost for cortical activation due to preferential use of glycolysis (Fox et al., 1986), as well as the concept of a large metabolic pool and small neurotransmitter/functional metabolic pool in neurons. As expressed in a review from that period (Raichle, 1998).

> *"These results suggested that the additional metabolic requirements associated with the increased neuronal activity might be supplied largely through glycolysis alone."*

Due to the low energy yield of glycolysis it was inferred that only a few percent of the energy supporting the brain was devoted to supporting neuronal signaling (see Hyder et al., 2013a for more recent calculations of the energy derived from glycolysis versus glucose oxidation). Therefore the large resting state imaging signals, whether CBF, $CMRO_2$, or CMRglc (which except under intense sensory activation is tightly coupled to $CMRO_2$), were not considered to reflect neuronal activity since they were primarily considered to be supporting the brain's non-signaling activities.

The concept of the majority of the brains energy not directly supporting function was challenged in 1998 with the finding, using ^{13}C MRS, that approximately 80% of cortical neuronal energy in the resting state was directly supporting neuronal signaling, as measured quantitatively with glutamate/glutamine cycling (Sibson et al., 1998). Subsequently this result has been replicated repeatedly and extended to GABAergic signaling and glial metabolism (Yu et al., 2017) in animals and humans as well as to independent electrical measurements of signaling (Hyder and Rothman, 2010; Hyder et al., 2011, 2013a). In addition multiple studies have shown that it is consistent with measurements of signaling and energetics at the cellular level when scaled up to whole cortex (Atwell and Laughlin, 2001; Yu et al., 2017). In 2001 Raichle et al. (2001) showed that PET measurements of the oxygen extraction fraction and fMRI measurements of deactivations were consistent with a high resting state level of neuronal activity relative to the fluctuations during tasks. In addition to providing localized signals during a task it was soon proposed that these changes were coordinated during the resting state by networks specialized for specific functions such as the default mode (Gusnard and Raichle, 2001). Similarly, Stark and Squire (2001) proposed high baseline activity as the explanation for paradoxical results found in neuroimaging studies of the mesial temporal lobe. In a recent review of imaging studies of brain function Raichle (2015) has proposed that the efficient brain use of the total energy supports signaling which clears up the previous questions on whether the presence of a high baseline signaling was regional (Raichle, 2010; Hyder et al., 2013a).

The high resting state neuronal activity (boXo in equation 1) (and associated neuroimaging signal would still not impact task based fMRI if the neuronal activity induced by a task or stimulus could be treated as being independent, as per the assumptions of linear models (Morcum and Fletcher, 2007). However, initially in animal models (Hyder et al., 2002; Zhu et al., 2009) and more recently in humans it has been shown that the increment or decrement measured during a task or stimulus is strongly dependent on the magnitude of the baseline neuronal activity (Uludag et al., 2004; Pasley et al., 2007; Hyder and Rothman, 2011). A similar dependence of the magnitude and pattern of fMRI fluctuations during the resting state upon the baseline neuronal activity measured by PET has recently been reported (Riedl et al., 2014; Aiello et al., 2015; Marchitelli et al., 2018).

Although there have been attempts to incorporate resting state activity into task/stimulus based functional neuroimaging there is no generally accepted procedure and the large majority of functional neuroimaging studies assume independence and either regress the baseline out (see Mortensen et al., 2018) or do not measure it as in fMRI. We note here that the global signal normalized out in resting state and sometimes task fMRI (Smith et al., 2013) is not a measure of the average resting state activity but rather correlated fluctuations in the low frequency fMRI difference signal across the entire cerebral cortex.

Non-linearity in the Imaging Response, Regional Interactions, and Networks

A key assumption in linear approaches is that the neuronal activity change induced within a voxel by a mental process (and the underlying signaling) is independent of the neuronal activity within the same voxel (and other voxels) supporting other mental processes. A range of studies have shown that this assumption can be violated, for example in studies in which there are competing

processes (Kastner et al., 1999; Reynolds and Desimone, 2003; Friston, 2009). Friston et al. (1996) have proposed that non-linear violations could be analyzed using a factorial approach in which the signal is described as the sum of regionally specific activations and regionally specific interactions between component mental processes. While this approach can better fit experimental data the assumptions of the validity of the cognitive models being used were not questioned, but rather how the brain supports those models:

"The point being made here is that although a cognitive science model, describing the functions, may include serial and additive elements the implementation of those functions is not. Consequently the structure of the cognitive components (functional model) and the brain's physiological implementation are not isomorphic and the mapping of one onto the other is problematic" (Friston et al., 1996).

More recently Price and Friston (2007) proposed a methodology by which neuroimaging could be used to help better define cognitive models using an approach they refer to as functional ontology, taking advantage of the interaction terms between neuronal activity in different brain regions. This and related work has been critiqued by Klein (2012) who argued that due to the many mappings between the same regions of the brain to different cognitive processes there is a need to include context dependence in the approach. While we feel these and related approaches to test and distinguish cognitive theories using neuroimaging data should be commended they differ from the NBM in that they largely focus on distinguishing cognitive theories rather than taking a complete bottom up approach in which theories of brain function are developed with functional neuroimaging and behavioral measurements, along with relevant neuroanatomical measurements, as the starting point.

Following from the work showing regional interactions network mapping has become an important area in functional neuroimaging both during tasks and at rest (Biswal et al., 1995; Hampson et al., 2006; Smith et al., 2013). However, in the majority of cases the networks identified are still assigned to supporting mental processes. Even for resting state fMRI the networks identified are usually related to a cognitive process such as the assignment of the default mode network (DMN) to the concept of general awareness/consciousness In Section "Application of the NBM to Studies Determining Neural Correlates of Consciousness," we describe studies in which the assignment of the DMN to a consciousness module was found to not correlate with the level of consciousness in coma patients as accurately as the NBM proposal, based on analysis of the total PET imaging signal during anesthesia, that total cortical activity was the best neuronal correlate of consciousness.

THE NON-COGNITIVE BEHAVIORAL MODEL (NBM)

The lack of a clear correspondence between the mental processes in models from cognitive science and anatomical localization of these processes by functional neuroimaging have led us to take a different approach, the NBM, in which cognitive models are not assumed from cognitive psychology but instead are empirically derived from imaging and other direct measurements of brain activity (Shulman et al., 2009, 2014; Shulman, 2013). In this section, we formalize this approach and contrast it with the SCM assumptions and methodology as applied to imaging.

Structure of the NBM

The NBM was developed from consideration of experimental results (Shulman et al., 2014) and from philosophical considerations on the psychological findings of Behaviorism (Shulman, 2013). We list the major assumptions and components of the model below.

Behaviorism as a Basis for NBM, the Absence of Assumed Mental Processes

Behaviorism is a psychology which proposes that behavior is a consequence or the effect of conditioning. Classical behaviorism was criticized for neglecting the state of the brain which is continuously engaging in internal behavior (for example the Freudian unconscious and in the cognitive psychology view the mental processes that compose SCM theories). We propose that due to functional neuroimaging it is possible to directly measure the internal activities of the brain, and therefore there is no need to introduce theoretical, abstract mental processes as an intermediate between inputs and behaviors. The role of experiment in NBM is to seek empirical connections between brain activity patterns developed by neuroplasticity in response to reproducible behavior or training as opposed to testing for brain regions that support predefined mental processes that act as an intermediate stage between brain activity and behavior.

This avoidance of mental concepts may be difficult to accept because these generalizations are of considerable value in daily life (e.g., concepts such as long term memory, attraction, intention, etc.) where they are firmly accepted in what is sometimes called "Folk Psychology." The effort to define what is meant by them and where they are supported in the brain has been a principle goal of neuroscientific and philosophical enquiry. However, in our opinion they have stood in the way of the development of novel data driven approaches to understanding how the brain supports behavior.

Neuroimaging Can Measure Patterns of Brain Activities Supporting Behavior

This is an inference from the many neuroimaging results that have shown that different behaviors are supported by different patterns of brain activity. However, it is an open question as to how reproducible these patterns are and how uniquely they map to behaviors. To the extent they are reproducible and generalizable across different behaviors it may be possible for experiments to identify the necessary patterns of activity needed, particularly for behaviors that have been reinforced by repetition (see the sections "fMRI Studies of Facial and Object Recognition" and "Application of the NBM to Study Cognitive Behaviors").

Measured Behavior Is Defined Operationally as Opposed to Being Based on Conceptual Generalizations

In SCM studies, the behavioral tasks studied are usually classified using concepts derived from cognitive psychology. In the NBM model, behavior is described only in terms of the actions performed. For example an NBM study of memory would expose the subject to all types of tasks that involve retrieving information. By contrast, in an SCM study the memory task would be classified according to differences in the cognitive processes they are hypothesized to contain such as working memory (and its sub components), long term memory and short term memory. This difference in parameters to be tested experimentally is the main operational difference between NBM and cognitive based models like SCM. Similarly, as described in Section "Application of the NBM to Studies Determining Neural Correlates of Consciousness," when NBM approaches were used to look for brain behavioral correlates of consciousness the state of consciousness was defined by how the subject responded to a standard list of simple questions about behavior from an experienced anesthesiologist (Shulman et al., 2009, 2014; Stender et al., 2015, 2016). This operational definition contrasts with the approach in most SCM neuroimaging studies of consciousness in which consciousness is defined in terms of cognitive concepts such as self awareness.

For Behaviors to Be Considered the Same or Similar They Require That the Associated Patterns of Neuronal Activity Be Similar

In the NBM behaviors are considered the same or similar only if both the measured behavior and the measured patterns of brain activity are the same or similar. For example if the brain pattern of activity differs between subjects performing the same task then the behavior is different even if the behavioral measurements are the same. In Section "Application of the NBM to Study Cognitive Behaviors," we argue that the poor reproducibility of studies trying to pinpoint the locations in the brain that support cognitive and other complex behaviors is due to the tasks being performed differently between subjects. The difference is due to the subject's different life experiences, and also due to feedback between behavior and brain during the study. In other words, the behavior itself is continuously modifying the patterns of brain activity supporting it due to neurofeedback. In SCM the behavior is usually described as in **Figure 1** as deriving directly from brain activity via functionally segregated regions supporting intermediate mental processes. In contrast in the NBM there is a continuous interdependence of brain activity and behavior as expressed by the back-and-forth arrows in **Figure 3**.

The Definition of Behavior Includes Internal Processes and Perception

Although not directly accessible by external behavioral measurements the NBM definition of behavior, including intrinsically internal processes such as silent reading, imagining, and object recognition, are considered as behaviors in the NBM. This inclusion differs from traditional behavioralism but is justified by imaging providing the ability to measure internal behavior.

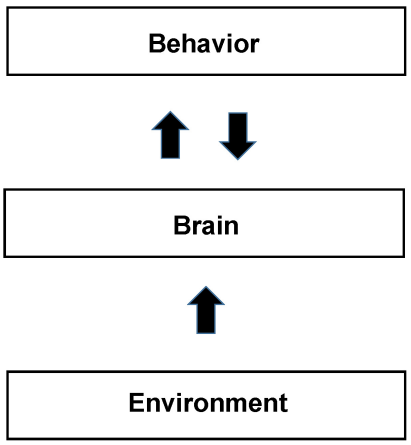


FIGURE 3 | Schematic of the Non-cognitive Behavioral Model (NBM). The figure shows a schematic diagram of the NBM. It differs from the SCM primarily by not assuming that the brain is organized to support localized cognitive and other separable mental processes. Brain support of behavior is derived from direct analysis of brain total activity maps from neuroimaging. The back and forth arrow between behavior and brain reflects their interdependence in addition to their separate causal factors. Identifications of both brain activity and behavior depend on their reproducibility and the precision of measurements. Behavior and brain activity are separated by measurements as opposed to the SCM view that behavior, arising from the mind, forms the relationship between mental and behavioral states. Sections "fMRI Studies of Facial and Object Recognition" and "Application of the NBM to Study Cognitive Behaviors" present examples that use behavior and brain measurements to understand whether apparently similar behavioral states are actually similar (as opposed to being so assumed from brain models). Neuroplasticity, in response to environmental experience, including training, develops readily available neural support mechanisms for individual behaviors. Due to the large variations in the experiences of individuals (see the section "Application of the NBM to Study Cognitive Behaviors") we expect significant intersubject differences in neural activity (and the resultant functional neuroimages), in the absence of reproducible behavior.

Context and Experience Plays a Key Role in How the Brain Supports Behavior Through Neuroplasticity and Other Low Level Mechanisms

In the SCM the components of high level cognitive processes are assumed to be generally present in all individuals, which justifies image and behavioral measurement averaging across individuals. This approach has been criticized for not taking context and individual experience into account (Shulman, 1996; Poldrack, 2006; Klein, 2012). The brain support of behavior depends upon context and individual experience acting through neuronal learning mechanisms. These mechanisms are shared between individuals but differences in their life experiences and specific task related training (e.g., object recognition as discussed in section "fMRI Studies of Facial and Object Recognition") may cause substantial differences in how the brain supports the behavior and even in the nature of the behavior itself to the subject.

Description of the Neuroimaging Signal in the NBM Formalism

In the NBM the neuronal activity in voxel i during behavior j (Nij) cannot be decomposed into separate linear contributions as in equation 1. If it is possible to describe the brain functionally based on the postulates of the NBM the signal in a voxel i during task j is described by a generalized function (eq. 5):

$$Nij = Fi(T + t) \quad (5)$$

Note that 'rest' is also considered a behavior (or state) so there is no separation between task and resting state imaging signals. The time variable T refers to the life history prior to the application of the task j in the study. The experiences acquired during the subject's life history influence the brain's pattern of neuronal activity supporting a behavior based upon neuroplasticity. In addition it determines the state of the brain at the time of the study (an extreme example of a state difference is a subject who is asleep in the scanner during the task. However, it is likely that much smaller state differences can significantly influence the brain's response).

Although, we have expressed the response here in terms of a function to clarify the differences with the SCM we do not know whether such a function actually can be found or even exists. Ultimately only experimentation and insights into brain organization and function at lower levels of description than the psychological level applied to most neuroimaging studies may be able to answer this question.

Experimental Application of the NBM

The NBM approach is in some ways conceptually similar to early electrical mapping and neuroimaging studies of Hubel (1988) and Zeki (1993) of the brain response to sensory stimulation. The identification of the functions supported by specific brain sub regions (e.g., columns in the primary visual cortex) was determined through trial and error in the exposure to a wide range of stimuli. While these studies identified microscopic level modularity (e.g., specific columns to edge detection, rotations, color detection, etc.) the functions of these regions were not assumed in advance, but were identified by experiment.

To apply this approach to cognitive neuroimaging requires that a range of behaviors be performed and the patterns of brain activity measured. We describe the result functionally in equation 6 to facilitate comparison with the SCM.

$$Nij = Fij(T + t) + e(t) \quad (6)$$

Behaviors (including internal behaviors such as perceptions) are not pre classified but rather tested for similarity based upon the consistency between the patterns of brain activity they induce. As described in Section "Application of the NBM to Study Cognitive Behaviors" this approach can potentially be used to determine which behaviors are related and how closely (through similarly of patterns of activity) as well as which regions are most directly involved in supporting a behavior The ability to statistically compare images without an underlying cognitive model does not present a novel challenge for imaging analysis. A wide range of methods already exist for testing pattern similarity in images (e.g., Kriegeskorte et al., 2008; Huth et al., 2016; Yourganov et al., 2014).

In order to illustrate more concretely the difference between the SCM approach we describe the application of the NBM to study working memory in comparison to that described in the Section "Basic Structure of Standard Cognitive Models (SCM) and Their Application to Functional Neuroimaging" and **Figure 2** as applied to the working memory model of Baddeley and Hitch (1974). The fundamental difference is that no cognitive process or processes for working memory would be assumed and used to design the memory tasks presented to the subject. Instead the subject would be presented different tasks in which memory would be operationally defined as involving performing recall or recognition of previously presented information (e.g., a list of words) in which parameters such as modality, time between presentation, task performance, and manipulation of presented information would be varied. The spatial and temporal pattern of the total neuronal activity (at a voxel level) would be measured for each memory task as well as the performance (behavior) of the subject. No normalization to take out resting state neuronal activity would be performed as in the SCM approach (the boXo term in eq. 1).

The results from the imaging of different memory tasks within a subject would then be compared using pattern analysis to assess the degree of similarity in the brain's response. As described in Section "Application of the NBM to Study Cognitive Behaviors" it may be possible to form classifications based upon the brain activity and behavior or tasks that are more or less similar. Based on these classifications, and assuming that results between subjects are sufficiently similar despite differences in their life history, it may be possible through further study to determine mechanisms (particularly through integration of results from different levels of study).

APPLICATION OF THE NBM TO STUDIES DETERMINING NEURAL CORRELATES OF CONSCIOUSNESS

There is an extensive history of studies attempting to identify neuronal correlates of consciousness (Kulli and Koch, 1991). As applied to neuroimaging these studies largely have focused on identifying the location of brain region(s) that support the mental processes responsible for the conscious state. This localization is based on the differential activity between these regions and the average activity of the total cortex, as expressed in terms of the PET CMRglc signal. In 2009, we analyzed previous experimental results from PET CMRglc studies of anesthesia using an early form of our NBM model (Shulman et al., 2009). We expand our analysis here using the NBM as described in this paper and apply it to new results that directly tested our conclusions.

The data initially analyzed were PET CMRglc images of human subjects at different levels of anesthesia and behavior. There have been multiple attempts, discussed below, to come up with an SCM based identification of specific brain regions, previously assigned to behavioral concepts like awareness, that support the state of consciousness. These studies identified the

role in consciousness contributed by these regions based on their showing a greater drop in the imaging signal during anesthesia than the average of the entire cerebral cortex. However, no consensus had been reached in these studies as to the anatomical localization of these cognitive components assigned to consciousness. By contrast an NBM analysis (Shulman et al., 2009) concluded that the best correlate with the behavior of a person in the state of consciousness was the total global cortical neuronal activity (as assessed from PET CMRglc). We describe below the differences between the SCM and NBM analysis to illustrate how the differences between the models adopted *ab initio* can lead to fundamentally different conclusions of how the brain supports behavior. We then describe a recent study of the clinically defined minimally conscious state (MCS) that directly compared the SCM and NBM predictions.

Although, we recognize that the studies of consciousness (often referred to as looking for correlates of consciousness) analyzed here do not follow the standard SCM experimental procedure in fMRI of within subject design they constitute a substantial body of research using both metabolic based imaging (fMRI, PET) and EEG/MEG based imaging. Furthermore they provide a clear test of the NBM proposal that the total imaging signal contains critical information on brain function and cannot be normalized away based on SCM assumptions regarding the separability of brain activity [see the section "Basic Structure of Standard Cognitive Models (SCM) and Their Application to Functional Neuroimaging"].

Description of Anesthesia Studies

Several laboratories (Alkire et al., 1995; Alkire, 2008; Katoh et al., 2000; Schlunzen et al., 2010) had performed PET neuroimaging studies on unconscious humans under surgical levels of anesthesia and in coma (Laureys et al., 2002, 2004) because of its clinical importance. In the anesthesia studies, subjects were studied both in the awake state and at various levels of anesthesia that reduced the average cortical (and sub cortical) energy consumption down to ∼50% of basal values, the level used for surgery. At different levels of anesthesia (and associated reduced CMRglc) subjects were assessed to be in a decreased state of consciousness based on their responses to questions and physical stimuli that are frequently used by anesthesiologists in clinical practice. Unconscious state cortical gray matter CMRglc was relatively uniform. Subsequent quantitative analysis has shown that the variations from the mean are no more than 10% (Hyder et al., 2013a).

Comparison of NBM and SCM Interpretations of the Anesthesia Results

Definitions of Consciousness

In the studies, where an SCM analysis was performed the behavior was defined in psychological terms in which consciousness was assumed to be a mental process that supported concepts such as self awareness. In contrast, in the NBM analysis the state of consciousness was identified by its behavior which was the ability of the subjects to respond to questions or physical stimuli from the anesthesiologist well enough to be classified as in the state of consciousness by standard metrics in the field. No attempt was made to relate the ability to respond to underlying cognitive descriptions of consciousness. The differences between the SCM conception of consciousness as a mental process conducted by functionally segregated brain regions and the NBM definition of it as a behavioral state of the subject altered the interpretation of the imaging results as described below.

Assignment of the Location of Consciousness Using the SCM and NBM Paradigms

Based on the SCM assumptions mental processes supporting consciousness were localized by looking for the regions that in the anesthetized state had the lowest level of FDG uptake relative to the average FDG uptake across the entire cortex. The large drop in global cerebral activity in all regions was assumed to not impact consciousness (and therefore was normalized away). This approach led to the brain support of the consciousness process being assigned, in different studies, to the thalamus, precuneus, inferior frontal lobe, and more recently the DMN (Alkire et al., 1995; Katoh et al., 2000; Laureys et al., 2004; Alkire, 2008). The variation in regional assignment is not surprising given that the regional fluctuations relative to the normalized global average imaging signal being at most 10% which is on the order of uncertainties in the imaging signal (Shulman et al., 2009; Hyder et al., 2013a).

In contrast, the NBM interpretation took into account the entire imaging signal since there was no *a priori* reason to distinguish the roles of different neuronal activities (i.e., total neuronal activity versus regional variations in activity). It was concluded (Shulman et al., 2009) that the averaged global drop in cerebral neuronal activity, as inferred from CMRglc PET measures of total energetics, was the best correlate of the state of consciousness. This drop was much larger than differences between the activity of different regions in either the conscious or unconscious states.

Statistical Comparison of SCM and NBM Predictions for Patients in the Minimally Conscious State and Vegetative State

The prediction from the NBM analysis that the global brain activity (as reflected in functional neuroimaging measurements of total glucose or oxygen consumption), would reflect the state of consciousness was recently tested in PET studies in a cohort of 41 patients in either a MCS or vegetative state (VS) as defined by standard neurophysiological measurements of Disorders of Consciousness (DOC) (Stender et al., 2015). Coma patients have long presented a challenge for predicting which patients will recover from the VS to reach consciousness and also whether the degree of consciousness of MCS patients can be enhanced.

In their study, Stender et al. (2015) found that the global CMRglc averaged 42% of control in the VS and 55% in the MCS. Regression analysis showed very little difference between values in several areas previously proposed to support consciousness processes, the brain stem, thalamus, precuneus, and frontal parietal cortex, and the global cortical activity, which explains the poor reproducibility of previous studies using the SCM

approach of looking for regions of maximum difference. Overall they concluded that total activity in the prefrontal lobe and the entire cortex were able to accurately distinguish states of consciousness. In a subsequent validation study that included 131 DOC patients (Stender et al., 2016). They reproduced these findings and established that 42% of normal cortical CMRglc is the minimal energetic requirement for conscious awareness.

Conclusions and Extension to the Study of Different States of Consciousness

In neuroimaging studies of anesthesia, and independently validated in coma, the identification of the total CMRglc signal (primarily reflecting neuronal signaling) by the NBM as the best correlate of consciousness was tested and confirmed in two studies of coma patients in different states of consciousness. The Stender studies found a somewhat better correlation with total cortical activity of the frontal parietal cortex, but the total CMRglc signal for both the global and frontal parietal cortex correlated better than differences in activity between regions previously assigned to selectively supporting consciousness.

A potential criticism of the NBM approach to interpreting and designing functional neuroimaging studies of consciousness is that it does not distinguish between different levels of consciousness and SCM derived properties, such as self-awareness, that are usually included in the definition of consciousness when physical correlates are looked for Crick and Koch (1990), Kulli and Koch (1991), Koch (2004). From an experimental standpoint, an NBM based study could be done by looking for correlations between images of brain activity (using a variety of imaging methods) and the level of response to any of a number of behavioral measures. Plots of the percent correctness of answering typical verbal questions at different levels of anesthesia (Katoh et al., 2000) have shown a non-linear but monotonic correlation of anesthesia with the variables responsible for the state of consciousness.

Although the global brain activity is the key parameter correlating with being in the state of conscious the NBM does not rule out that patterns of brain activity may vary at different levels of conscious activity. In their 2015 paper, Stendor et al. found that the standard deviation in the total cortical activity increased as a percentage of the global average at higher states of consciousness. This finding is consistent with MCS patients often having the full range of sensory and cognitive abilities (e.g., language) although at a lower functioning level supporting a key role for global cortical activity in specific behaviors (which in the SCM is treated as independent of cortical activity).

In addition to resolving the questions evoked by the term consciousness the NBM framing of research questions in terms of observable behaviors and the identification of an experimental parameter, the total global energy, which, quantitatively measured, opens the door to novel experimental studies. By allowing experimental results to define the parameters of interest instead of insisting on fitting a cognitive model the NBM promises to continue to provide novel insights into consciousness. The important novelty of our approach is that we identify a brain property when the person is in the behavioral state of consciousness that changes when he or she leaves that state. This along with analyzing the total brain activity within brain regions has allowed a stronger correlate of consciousness to be identified than the SCM interpretation which assumes that a localized modular response varies with conscious behavior. Future studies of the relationships between the global and regional levels of brain activity at different levels of consciousness (Kotoh) may reveal further insight into their interplay in supporting specific conscious behaviors.

fMRI STUDIES OF FACIAL AND OBJECT RECOGNITION

In this section, we assess whether the SCM assignment of the fusiform gyrus (FG) region as supporting a facial recognition process or module is a better explanation than the training and neuroplasticity proposed in the NBM. Although these studies were not performed using the NBM paradigm we believe they provide important support for the importance of neuroplasticity and personal experience in developing the brain's support of behavior.

The fusiform gyrus is a brain region intermediate between visual and cognitive processing that has shown good image reproducibility. Early functional imaging studies had identified a facial recognition region in the FG by the difference between the neuronal signal from faces right-side up and up-side down (Allison et al., 1994; Puce et al., 1996). The differences in the FG between up and down orientations of human faces gave a reproducible fMRI difference signal leading to this locale being defined as the fusiform face area (FFA) region. The ease of finding this response to faces led to the assignment of this region of the FG as "the FFA: a module in human extrastriate cortex specialized for face perception" (Kanwisher et al., 1997). However, differing from this interpretation, other studies showed that the FFA responses that experts have learned to familiar objects (e.g., of bird watchers to birds) supported a similar reproducible response to familiar "non-face objects." The similar responses in the FFA of untrained persons to faces and of experts to familiar objects have been interpreted either as the activations of innate modules, consistent with the SCM (Kanwisher et al., 1997) or as the effects of expertise developed from training, consistent with the NBM, in which the FFA is part of a network tuned by experience to individuate visually similar objects. Experiments were performed that distinguish these two interpretations which we review here.

Based on the convention of researchers in the field (e.g., Kanwisher et al., 1997) we refer to the mental process of facial recognition in this section as a module. The term module based on historic usage (see Fodor, 1983, 2000) refers to a mental process that is almost completely isolated from other processes (other than inputs and outputs) and is anticipated to have a similar degree of isolation in its implementation by the brain. Supporting this assumption for the fusiform gyrus, which is part of the visual processing stream, a high (although not complete) degree of modularity has been found in regions of the visual cortex particularly V1.

fMRI Studies of the Effect of Training on the FFA Response

To clarify the experimental distinction between the two brain models, studies aimed to distinguish whether the enhanced neural response in the FG was due to an innate module for facial recognition or due to the expertise created by training. Among the early studies addressing this issue were fMRI experiments by Gauthier et al. (1999) who compared the response to faces with that to face-like objects called Greebles. In human subjects the differences in fMRI images between up and down orientations of a human face had given a reproducible difference signal in the fusiform gyrus, leading to its assignment as a FFA The Greebles were designed to differ from each other, with individuals falling into classes based upon more prominent jaws and smaller nose-regions. Subjects were tested on their ability to recognize specific Greebles with differences between the ability of trained and untrained subjects to distinguish between the right-side up and upside down Greebles as a control. They measured fMRI difference signals from human subjects trained to recognize Greebles vs. signals from untrained controls to ascertain the degree of expertise and then measured the difference signal by subtracting the fMRI images from an upside down Greeble from one when it was right side up. They hypothesized that if an FFA-like activation arose with the development of expertise it would be evidence that the region is not innately specialized to recognize Greebles but rather is recruited through plasticity and experience.

In their original fMRI study (Gauthier et al., 1999) of untrained humans, no reproducible differences were observed in the FFA between the two orientations of the Greeble. However, after subjects had been trained to recognize Greebles, the different orientations of the Greebles gave fusiform gyrus activation quite similar to those raised by faces. Additional experiments to measure the difference signal from experts in identifying other familiar objects, e.g., animals, automobiles and planes also found activated regions in the FFA (Gauthier et al., 2000; Xu, 2005) The various objects, including Greebles, whose training history was definitely known, activated the same FFA region in experts as faces, leading to the conclusion that the response to facial recognition was not different from the response from experts who had been trained to recognize the objects. Therefore the postulation of an innate facial recognition region, different from the attributes of expertise, was not supported by the data which showed that the response to faces was similar to the response from experts trained to recognize familiar objects (Gauthier et al., 1999).

Ultra High Resolution Studies of the FFA Provide Further Support of a Non-modular Interpretation

The findings of Gauthier et al. (1999) and others denying the modular nature of object recognition were criticized (Tsao et al., 2006) based on the resolution of the fMRI studies being 2 mm × 2 mm × 3 mm which, it was claimed, could lead to regions specialized for only facial recognition overlapping with other regions. To test this possibility studies were performed by at higher spatial resolution (approximately 1 mm^2 in plane). The initial high resolution studies provided ambiguous results due to low signal to noise (Grill-Spector et al., 2006). However, a subsequent study at 7T, which provided higher sensitivity, showed conclusively that the same FFA region was activated by both facial recognition and by an expert's recognition of familiar objects (McGugin et al., 2012). Hence, since the recognition of familiar objects like automobiles, birds and Greebles, where expertise was developed by training, were identical to the signals from faces, the simplest conclusion was that faces, along with these other familiar objects, led to the development of a region in the FG of increased brain activity via training, not by an innate FFA module existing for each of the familiar objects. A further argument against an innate FFA is that cognitive processes specializing in objects only in existence over the last two centuries such as automobiles and planes could not have been selected by evolution.

Localized Neuroimaging Responses Do Not Imply Cognitive Processes: An Alternate Explanation of Imaging Evidence for Functional Segregation Based Upon Neuroplasticity and Image Averaging

Localized regions of activation in neuroimages, as found in the FG studies, are often cited as evidence for underlying separable cognitive processes. However, detailed examination of the FG studies, discussed above, where well localized imaging responses, albeit subject to the limitations described in Sections "Basic Structure of Standard Cognitive Models (SCM) and Their Application to Functional Neuroimaging" and "The Non-cognitive Behavioral Model (NBM)" of SCM experimental paradigms, are found when subjects have expertise in recognition, have been shown to not support the presence of an innate modular FFA. Instead localization and strength of the neuroimaging response depend on expertise and training. The question, however, remains regarding how do well localized regions of enhanced (or decreased) neuronal activity arise in an image.

Due to limitations in sensitivity for both PET and fMRI the images obtained during tasks, particularly cognitive tasks, must be repeated multiple times and added together. Often the results from several subjects are added. As such, a reproducible pattern of neuronal activity during a behavior will create a greater average neuroimaging signal than less reproducible patterns even if they are supported by similar overall amounts of neuronal activity. The NBM hypothesizes that a reproducibile pattern of brain activity in response to an input, results from repetition leading to the pattern being selected for using brain mechanisms for neuroplasticity.

Conclusions

Neuroimaging studies of object and facial recognition show a dependence on expertise consistent with the NBM interpretation that neuroplasticity and experience play a key role in how the brain supports behavior. Expertise comes from recruitment of regions within the FG (and other brain regions) that

depend upon training and experience as opposed to being an innate module. The ubiquity of facial recognition without any training presumably reflected the many exposures to faces at an early date thereby training people so that the fMRI data are thereby consistent with the expectations of the NBM. In Section "Application of the NBM to Study Cognitive Behaviors," we discuss how reproducible neuronal activity (and the neuroimaging signals it is mapped by) develops through experience and why sensory processing regions show much higher reproducibility than cognitive regions.

We caution that the agreement with experiment depends upon selecting results that, while they seem valid to us, must be recognized as not uncontested. In addition all of the studies described above were performed using the SCM framework and were therefore designed to identify the location(s) of a previously hypothesized mental process, whether from an innate modular type structure or a similar structure derived from experience. Within the framework of this paper, which intends to describe the NBM model and its advantages in interpreting specific experimental results, we cannot claim that our presentation offers a balanced review of the many results available from relevant brain imaging experiments. However, by concentrating on experiments that have intended to distinguish between modular and non-modular interpretations we hope to have clarified the nature of the disagreement between SCM and NBM and to show that the NBM interpretation of the Greeble study as supporting reproducible training being the origin of FFA localized brain activity.

APPLICATION OF THE NBM TO STUDY COGNITIVE BEHAVIORS

In this section, we further address the question of how NBM and other bottom up models can be used to study cognition and other complex behaviors. This question was highlighted soon after the discovery of fMRI in an interview with one of the authors (Shulman, 1996), and is largely based upon the lack of a direct theory of cognition (such as is embedded in the SCM) being tested (Morcum and Fletcher, 2007). We describe here a program for studying the brain/behavior relationship that uses neuroimaging to identify neuronal mechanisms supporting cognition based upon the similarity of patterns of brain activity and directly taking into account context and the experiences of the subjects.

The NBM as an Explanation for the Lack of Unique Functional Segregation for Supporting Different Cognitive Processes

As described in Section "Basic Structure of Standard Cognitive Models (SCM) and Their Application to Functional Neuroimaging" neuroimaging studies using the SCM paradigm have not found clear functional segregation of cognitive processes which has been attributed to limitations in imaging reliability as well as non-linearities and regional interactions in the brain's implementation of cognitive processes (Friston et al., 1996; Price and Friston, 2007; Gonzalez-Castillo et al., 2012). In the NBM, we propose that the lack of clear functional segregation is a consequence of the brain not being organized to support abstract cognitive processes as described by the SCM. Instead the specific neuronal instantiations of behavior depends largely upon training and experience acting upon mechanisms of neuroplasticity. Due to the large variations in the experiences of individuals patterns of neuronal activity supporting behaviors will also exhibit large variations. Furthermore differences in an individual's history changes the context in which a behavior is performed or interpreted. For example, the meaning of concepts like patriotism, memory or beauty will vary considerably between individuals, with different histories, so that the brain activity induced by the same word will depend upon its context.

The higher reproducibility of the neuronal response to sensory stimuli is driven by the reproducible training in sensory tasks generated in everyday life, i.e., we all can accurately distinguish red from blue and up from down. The specific brain responses learned by early repetitive exposure to sensory stimuli show high functional segregation in that they become sensitized to a similar visual feature, e.g., moving lines or edges or red versus blue. The easy transferability of specific sensory phenomena relative to mental concepts is due to their being reproducibly defined by environmental input. Psychological concepts like working memory are not reproducible because they are not uniformly defined between individuals nor in the same individual in different contexts[2].

How Mental Processes Can Be Studied With the NBM

In NBM there is no equivalent to the traditional SCM interpretation of neuroimages in which the brain is functionally organized to support predefined mental processes. Instead similarities in the pattern of brain activity determines whether behaviors are related. For example suppose one wants, from an NBM guided experiment, to know the brain activities needed during a set of behaviors that can be described in somewhat general behavioral terms as working memory. In the NBM there is no definition of a concept of working memory that deploys the same mental process in different behaviors. The behaviors could be described as including something that could be called working memory only if you were willing to sacrifice accuracy to obtain such a generalization and were willing to overlook differences in the brain activity supporting the behaviors.

[2]In contrast to the human studies, studies of non-human primates performing cognitive tasks have been shown to give highly reproducible localized patterns of activity after they were strongly trained to perform such tasks. These observations (e.g., Goldman-Rakic, 1995) have encouraged neuroscientists to believe that cognitive concepts like WM are represented in the brain in a neuroanatomically reproducibile and well defined manner, despite functional neuroimaging studies in humans showing poor localization (see the section "Basic Structure of Standard Cognitive Models (SCM) and Their Application to Functional Neuroimaging" and references therein). Our alternate interpretation is that observation of the expected brain activities in non-human primates after such intense training, but not in un-trained human subjects, argues against intrinsic functional segregation of brain activity supporting the cognitive process of working memory. Instead the functional segregation derives from repeated training and eventual selection of a fixed neuronal pathway for supporting the behavior.

In the NBM paradigm, behaviors traditionally assigned to concepts such as memory would be studied using the same approach as the studies of consciousness and object/face recognition in Sections "Application of the NBM to Studies Determining Neural Correlates of Consciousness" and "fMRI Studies of Facial and Object Recognition." In Section "fMRI Studies of Facial and Object Recognition," it was determined from the images that the brain activity supporting expert object recognition and face recognition in the FG were similar, implying related neural mechanisms. Similarly, studies of "memory" would involve having the subject perform a series of behaviors that involve memory defined operationally as the ability to respond to questions about previous knowledge. However, any conclusion about whether the brain responses to the behaviors so defined as memory are similar would depend upon neuroimaging establishing that they use related neuronal mechanisms. The problem that might remain is whether the brain implementation has sufficient similarities between individuals to identify common mechanisms. It might be that this definition of "memory" might differ between individuals, given the same definition of behavior, like the synonyms in a dictionary that invariably accompany the definition of a term. In the SCM, mental processes often are produced by words that encompass their activity but it is unclear if the concepts embodied in words will be distinguished at the level of neural mechanisms. The breakdown of memory in SCM guided experiments into working memory and other kinds of memory suggests that different kinds of kinds of memory are distinguished and future work might be able to identify more useful distinctions while acknowledging their uniqueness as a consequence of the life experiences (external and internal) of the individual.

A potential limitation of the NBM approach is that the variation within and between subjects will be too high in order to find correlations, particularly given the limited signal to noise of fMRI. The large majority of studies have been performed using SCM paradigms in which there are significant constraints on the data analysis in order to achieve statistical significance, albeit at the cost of potentially forcing results to agree (see the section "Basic Structure of Standard Cognitive Models (SCM) and Their Application to Functional Neuroimaging" and references cited within). However, with improvements in fMRI sensitivity it has been possible to do relatively unconstrained analysis of the pattern of brain response to visual tasks and find reproducibility at least within subjects (Gonzalez-Castillo et al., 2012) as well as in resting state fMRI (Smith et al., 2013) networks that contain regions that have been assigned (albeit by an SCM approach) to cognition. Ultimately, however, experiments will need to be performed in order to answer the question of whether common mechanisms can be identified using an assumption free approach like the NBM.

The NBM in the Study of Consciousness

The SCM goal was to find a brain region activated during the cognitive acts defined as consciousness. This goal has failed, because of the SCM requirements first to find agreement on the definition of "consciousness" and second to identify the brain regions activated. These failings have been avoided by planning and interpreting the search by NBM. In this approach the behavior is identified by observing when the subject in the state of consciousness while the brain region to be associated with that behavior need not be localized. We see here two advantages of the NBM method. To study consciousness by an SCM type experiment means locating it in a brain region, which also requires that it is necessary to agree about the definition of the term. Since neuroscience has been severely criticized by philosophers and others for not being able to define consciousness (Nagel, 2016) efforts by studies using the SCM to study consciousness would only succeed if first there had been such a consensus and second if it resulted in well defined functional segregation. By contrast using the NBM experimental approach of direct statistical correlation it was found that the strongest predictor of consciousness was the total, global, energy consumption by the brain when the person was identified by behavior as being in the state of consciousness.

Conclusion

In conclusion, we propose that the brain's support of cognitive behavior is highly experience-dependent and sensitive to the context of the study. As people have normally been trained during their early years to recognize sensory stimuli their brains will show similar patterns of neuronal activity when the same stimuli are presented, and when summed will give reproducible and well localized neuroimaging activity maps. Furthermore once trained by such reproducible exposures the brain activity will continue to be activated by the stimulus. However, as generalized behaviors like remembering, calculating or paying attention depend on their context and on the person's history, the neuronal activity needed to support them will vary and generally will not yield highly reproducible functional neuroimages. Using the neuroimaging activity maps as a guide, it is proposed by NBM to identify behaviors that are supported by related brain mechanisms. However, it is unlikely, based on present results, that the degree of reproducibility required to support models will be found in cognitive modules and therefore it is necessary to look for empirical mechanisms to tighten the correlation between brain activity and behavior.

EPISTEMOLOGICAL BASIS OF THE NBM

To the degree that NBM can find reliable relations between observable behavior and brain patterns of neuronal activity, we have fulfilled a goal of our modified behaviorism. Therefore, we stop to ask what have we achieved? Does NBM have any generalized importance beyond the knowledge that a certain brain region responds to a certain person's behavior? Does it tell us anything of general usefulness or is it limited to the particular experimental conditions from which we drew correlations? This is the recurring question faced by all models of brain function that are not based on conceptualizations of mental processes attributed to behavior such as are offered by the SCM.

One of the most significant contributions to this problem was offered by Charles Sanders Peirce in his extension of

philosophical pragmatism (Peirce, 1958). Peirce suggested that a unifying evaluation of such differing results could be found by considering the consequences of these separate but similar relationships upon human actions. In his method for identifying the meaning of a pragmatist conclusion, Peirce proposed a criterion for understanding the consequences of assigning a generalized conceptualization to a well-defined behavior, which we have extended to its correlation with a brain activation. For Peirce the value of concepts behind behaviors like working memory, free will or unselfishness was not to be determined by how accurately they described the phenomenon, but rather by the meaning they had for human affairs. His definition was intended to return the term to scientific purview by defining it in terms of scientific "thought" experiments as follows- "Consider, what effects, that might conceivably have practical bearings, we conceive the object of our conception to have. Then, our conception of these effects is the whole of our conception of the object" (Ibid, p. 192). This includes the likelihood that a brain response for a somewhat generalized act will include different brain responses reflecting the usage allowed by the several synonyms of the term in a dictionary. Since the meaning of a concept depends on the effects it may have on humans, this definition has not assigned a value to this meaning but has identified its usage in a common expression.

As discussed by Craver (2007), we need some principled and empirical way of saying when observable behaviors can be correlated without previous assumptions of similarity. It is here that Peirce's turning to the possible effects of the understanding of the behavior upon human actions, as the criterion of meaning, rescues us from the criticism of having found only a subjective result. In decomposing the behavior into little behavioral steps, not into cognitive concepts, we might find linked brain correlates of the jump, in brain neuronal activity. Furthermore unlike the SCM there is not a starting assumption about mental processes the brain is performing, instead there are details of behavior that can be teased out which are ignored by assuming classes of behaviors are similar. Furthermore by relating detailed behaviors to the effects upon human activity following Peirce or upon the correlated brain activity as proposed in NBM we would be on the way to coordinating brain activations during a behavior, a fundamental goal of neuroscience.

Although our proposal of an empirically based model of brain function has not, to our knowledge, been previously generalized from neuroimaging data, still specific, similar empirical models have been proposed (Koch, 2017) by authors who previously had leaned toward interpreting consciousness by modular-like concepts (Koch, 2004). This move, linking conscious behavior to a measurement of its electrical activity formally resembles our NBM model rather than the many former studies of consciousness that searched for a brain activity linked to a cognitive concept. Tononi et al. (2016) propose that measuring brain electrical activity after stimulating the brain with magnetic pulses provides a reliable measurement of consciousness. They proposed that an empirical level of brain activity, which they obtained from comparison of conscious and unconscious subjects, could be used to define consciousness.

In substituting correlations for causal relationships between observed behavior and brain activities, Koch and Tononi's model, similar to our NBM and to Peirce's proposal for future experiments resembles the recent progress in machine learning to find relationships between data and behavior without prior hypotheses. The vast amount of data obtained even in a single functional neuroimaging study in principle would be well suited to this type of analysis. In addition this approach could be integrated with models of brain function derived from bottom up lower level models of brain circuitry and general informational principles. Significant attempts have been made in these areas that potentially could be integrated into modeling neuroimaging studies (e.g., Bedau, 1997; van der Maas et al., 2006; Kitzbichler et al., 2009; Hellyer et al., 2017). By taking advantage of the developing methods of machine learning our empirical approach to neuroimaging provides an exciting future for understanding brain functions.

AUTHOR CONTRIBUTIONS

RGS and DLR contributed to the writing of the manuscript, the hypotheses and theories presented, and the analysis of data reviewed.

FUNDING

DLR was supported by grants from the National Institute of Health R01NS100106 and R01NS087568A.

ACKNOWLEDGMENTS

We thank Carl Craver for his invaluable assistance in helping to define the differences between the SCM and NBM models and clarifying other points made in the manuscript.

REFERENCES

Aiello, M., Salvatore, E., Cachia, A., Pappatà, S., Cavaliere, C., Prinster, A., et al. (2015). Relationship between simultaneously acquired resting-state cerebral glucose metabolism and functional MRI: a PET/MR hybrid scnner study. *Neuroimage* 113, 111–112. doi: 10.1016/j.neuroimage.2015.03.017

Alkire, M. T. (2008). Probing the mind: anesthesia and neuroimaging. *Clin. Pharmacol. Ther.* 84, 149–152. doi: 10.1038/clpt.2008.75

Alkire, M. T., Haier, R. J., Barker, S. J., Shah, N. K., Wu, J. C., Kao, Y. J., et al. (1995). Cerebral metabolism during propofol anesthesia in humans studied with positron emission tomography. *Anesthesiology* 82, 393–403. doi: 10.1097/00000542-199502000-00010

Allison, T., Ginger, H., McCarthy, G., Nobre, A. C., Puce, A., Luby, M., et al. (1994). Face recognition in human extrastriate cortex. *J. Neurophysiol.* 71, 821–826. doi: 10.1152/jn.1994.71.2.821

Ashburner, J. (2012). SPM: A history. *Neuroimage* 62, 791–800. doi: 10.1016/j.neuroimage.2011.10.025

Atwell, D., and Laughlin, S. B. (2001). AN energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* 21, 1133–1145. doi: 10.1097/00004647-200110000-00001

Baddeley, A. D., and Hitch, G. (1974). *Working Memory*. New York, NY: Academic Press.

Bedau, M. A. (1997). "Weak emergence," in *Philosophical Perspectives: Mind, Causation and World*, ed. J. Tomberlin (Malden, MA: Blackwell), 375–399.

Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo- planar MRI. *Magn. Reson. Med.* 34, 537–541. doi: 10.1002/mrm.1910340409

Chomsky, N. (1959). A review of B. F. Skinner's verbal behavior *Language* 35, 26–58. doi: 10.2307/411334

Chomsky, N. (1971). *The Case Against B.F. Skinner*. New York City, NY: The New York Review of Books.

Cosmides, N. (1980). *Rules and Reprentations*. New York, NY: Columbia University Press.

Craver, C. F. (2007). *Explaining the Brain- Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199299317.001.0001

Creutzfeldt, O. D. (1975). "Neurophysiological correlates of different functional states of the brain," in *Alfred Benzon Symposium VII*, eds D. H. Ingvar and N. A. Lassen (Cambridge, MA: Academic Press), 21–46.

Crick, F., and Koch, C. (1990). Some reflections on visual awareness. *Cold Spring Harb. Symp. Quant. Biol.* 55, 953–962. doi: 10.1101/SQB.1990.055.01.089

Donders, F. C. (1969). The speed of mental processes. *Acta Psychol.* 30, 412–431. doi: 10.1016/0001-6918(69)90065-1

Fodor, J. (1983). *The Modularity of Mind*. Cambridge, MA: The MIT Press.

Fodor, J. (2000). *The Mind Doesn't Work That Way: The Scope And Limits Of Computational Psychology*. Cambridge, MA: The MIT Press. doi: 10.7551/mitpress/4627.001.0001

Fox, P. T., Mintun, M. A., Raichle, M. E., Miezin, F. M., Allman, J. M., and Van Essen, D. C. (1986). Mapping human visual cortex with positron emission tomography. *Nature* 323, 806–809. doi: 10.1038/323806a0

Frackowiak, R. S. J., Ashburner, J. T., Penny, W. D., Zeki, S., Friston, K. J., Frith, C. D., et al. (2004). *Human Brain Function*, 2nd Edn. San Diego, CA: Academic Press.

Friston, K. J. (1997). "Analyzing brain imagies:Principles and overview," in *Human Brain Function*, eds R. S. J. Frackowiak, K. J. Friston, C. D. Frith, R. J. Dolan, and J. C. Maziotta (Cambridge, MA: Academic Press), 25–41.

Friston, K. J. (2005). Models of brain function in neuroimaging. *Annu. Rev. Psychol.* 56, 57–87. doi: 10.1146/annurev.psych.56.091103.070311

Friston, K. J. (2009). Modalities, modes, and models in functional neuroimaging. *Science* 326, 399–403. doi: 10.1126/science.1174521

Friston, K. J., and Price, C. J. (2011). Modules and brain mapping. *Cogn. Neuropsychol.* 28, 241–250. doi: 10.1080/02643294.2011.558835

Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S., Dolan, R. J., et al. (1996). The trouble with cognitive subtraction. *Neuroimage* 4, 97–104. doi: 10.1006/nimg.1996.0033

Gauthier, I., Skudlarski, P., Gore, J. C., and Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nat. Neurosci.* 3, 191–197. doi: 10.1038/72140

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., and Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat Neurosci.* 2, 568–573. doi: 10.1038/9224

Gazzaniga, M. S., and Mangun, G. R. (eds) (2014). *The Cognitive Neurosciences*, 5th Edn. Cambridge, MA: MIT Press.

Goldman-Rakic, P. M. (1995). Cellular basis of working memory. *Neuron* 14, 477–485. doi: 10.1016/0896-6273(95)90304-6

Gonzalez-Castillo, J., Saad, Z. S., Handwerker, D. A., Inati, S. J., Brenowitz, N., Bandettini, P. A., et al. (2012). Whole-brain, time-locked activation with simple tasks revealed using massive averaging and model-free analysis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 5487–5549. doi: 10.1073/pnas.1121049109

Grill-Spector, K., Sayres, R., and Ress, D. (2006). High resolution imaging reveals highly selective non-face clusters in the fusiform face area. *Nat. Neurosci.* 9, 1177–1185. doi: 10.1038/nn1745

Gusnard, D., and Raichle, M. E. (2001). Searching for a baseline:functional imaging and the resting brain. *Nat. Rev. Neurosci.* 2, 685–694. doi: 10.1038/35094500

Hampson, M., Driesen, N. R., Skudlarski, P., Gore, J. C., and Constable, R. T. (2006). Brain connectivity related to working memory performance. *J. Neurosci.* 26, 13338–13343. doi: 10.1523/JNEUROSCI.3408-06.2006

Hellyer, P. J., Clopath, C., Kehagia, A. A., Turkheimer, F. E., and Leech, R. (2017). From homeostasis to behavior: balanced activity in an exploration of embodied dynamic environmental-neural interaction. *PLoS Comput. Biol.* 13:e1005721. doi: 10.1371/journal.pcbi.1005721

Hoge, R. D., Atkinson, J., Gill, B., Crelier, G. R., Marrett, S., Pike, G. B., et al. (1999). Linear coupling between cerebral blood flow, and oxygen consumption in activated human cortex. *Proc. Natl. Acad. Sci. U.S.A.* 96, 9403–9408. doi: 10.1073/pnas.96.16.9403

Hubel, D. H. (1988). *Eye, Brain and Vision*. New York, NY: Scientific American Library.

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., and Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 532, 433–458. doi: 10.1038/nature17637

Hyder, F., Fulbright, R. K., Shulman, R. G., and Rothman, D. L. (2013a). Glutamatergic function in the resting awake human brain is supported by uniformly high oxidative energy. *J. Cereb. Blood Flow Metab.* 26, 865–877. doi: 10.1038/jcbfm.2012.207

Hyder, F., Rothman, D. L., and Bennett, R. (2013b). Cortical energy demands of signaling and nonsignaling components in brain are conserved across mammalian species and activity levels. *Proc. Natl. Acad. Sci. U.S.A.* 110, 3549–3554. doi: 10.1073/pnas.1214912110

Hyder, F., Herman, P., Sanganahalli, B. G., Coman, D., Blumenfeld, H., Rothman, D. L., et al. (2011). Role of ongoing, intrinsic activity of neuronal populations for quantitative neuroimaging of functional magnetic resonance imaging-based networks. *Brain Connect.* 1, 185–193. doi: 10.1089/brain.2011.0032

Hyder, F., and Rothman, D. L. (2010). Neuronal correlate of BOLD signal fluctuations at rest: err on the side of the baseline. *Proc. Natl. Acad. Sci. U.S.A.* 2010, 10773–10774. doi: 10.1073/pnas.1005135107

Hyder, F., and Rothman, D. L. (2011). Evidence for the importance of measuring total brain activity in neuroimaging. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5475–5476. doi: 10.1073/pnas.1102026108

Hyder, F., and Rothman, D. L. (2012). Quantitative fMRI and oxidative neuroenergetics. *Neuroimage* 62, 985–994. doi: 10.1016/j.neuroimage.2012.04.027

Hyder, F., Rothman, D. L., and Shulman, R. G. (2002). From the Cover: total neuroenergetics support localized brain activity: Implications for the interpretation of fMRI. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10771–10776. doi: 10.1073/pnas.132272299

Jonides, J., Lewis, R. I., Nee, D., Lustig, C. A., Berman, M. G., and Moore, K. S. (2008). The mind and brain of short time memory. *Ann. Rev. Psychol.* 59, 193–224. doi: 10.1146/annurev.psych.59.103006.093615

Kanwisher, N., McDermott, J., and Chun, M. M. J. (1997). The fusiform face area: a model in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311. doi: 10.1523/JNEUROSCI.17-11-04302.1997

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., and Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron* 22, 751–761. doi: 10.1016/S0896-6273(00)80734-5

Katoh, T., Bito, H., and Sato, S. (2000). Influence of age on hypnotic requirement, bispectral index, and 95% spectral edge frequency associated with sedation induced by sevoflurane. *Anesthesiology* 292, 55–61. doi: 10.1097/00000542-200001000-00014

Kitzbichler, M. G., Smith, M. L., Christensen, S. R., and Bullmore, E. (2009). Broadband criticality of human brain network synchronization. *PLoS Comput. Biol.* 5:e1000314. doi: 10.1371/journal.pcbi.1000314

Klein, C. (2012). Cognitive ontology and region versus network oriented analysis. *Philos. Sci.* 79, 952–960. doi: 10.1086/667843

Koch, C. (2004). *The Quest for Consciousness: A Neurobiological Approach*. Englewood, CO: Roberts & Co.

Koch, C. (2017). How to make a consciousness meter. *Sci. Am.* 317, 28–33. doi: 10.1038/scientificamerican1117-28

Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2:4. doi: 10.3389/neuro.06.004.2008

Kulli, J., and Koch, C. (1991). Does anesthesia cause loss of consciousness? *Trends Neurosci.* 14, 6–10.

Laureys, S., Antoine, S., Boly, M., Elincx, S., Faymonville, M. E., Berre, J., et al. (2002). Brain function in the vegetative state. *Acta Neurol. Belg.* 102, 177–185.

Laureys, S., Owen, A. M., and Schiff, N. D. (2004). Brain function in coma, vegetative state, and related disorders. *Lancet Neurol.* 3, 537–546. doi: 10.1016/S1474-4422(04)00852-X

Marchitelli, R., Aiello, M., Cachia, A., Quarantelli, M., Cavaliere, C., Postiglione, A., et al. (2018). simultaneous resting-state FDG-PET/fMRI in Alzheimer disease:relationship between glucose metabolism and intrinsic activity. *Neuroimage* 176, 246–258. doi: 10.1016/j.neuroimage.2018.04.048

McGonigle, D. J., Howseman, A. M., Athwal, B. S., Friston, K. J., Frackowiak, R. S., Holmes, A. P., et al. (2000). Variability in fMRI: an examination of intersession differences. *Neuroimage* 11, 708–734. doi: 10.1006/nimg.2000.0562

McGugin, R. W., Gatenby, J. C., Gore, J. C., and Gauthier, I. (2012). High Resolution imaging of expertise reveals reliable object selectivity in the fusiform face area related to perceptual performance. *PNAS* 109, 17063–17068. doi: 10.1073/pnas.1116333109

Morcum, A. M., and Fletcher, P. C. (2007). Does the brain have a baseline? Why we should be resisting a rest. *Neuroimage* 37, 1073–1082. doi: 10.1016/j.neuroimage.2006.09.013

Mortensen, K. N., Gjedde, A., Thompson, G. J., Herman, P., Parent, M. J., Rothman, D. L., et al. (2018). Impact of global mean normalization on regional glucose metabolism in the human brain. *Neural Plast.* 2018:6120925. doi: 10.1155/2018/6120925

Nagel, T. (2016). An exchange with thomas naggel: the mind-body problem and psychoanalysis. *J. Am. Psychoanal. Assoc.* 64, 389–403. doi: 10.1177/0003065116647053

Otzenberger, H., Gounot, D., Marrer, C., Namer, I. J., and Metz-Lutz, M. N. (2005). Reliability of individual functional MRI brain mapping of language. *Neuropsychology* 19, 484–493. doi: 10.1037/0894-4105.19.4.484

Pasley, B. N., Inglis, B. A., and Freeman, R. D. (2007). Analysis of oxygen metabolism implies a neural origin for the negative BOLD response in human visual cortex. *Neuroimage* 36, 269–276. doi: 10.1016/j.neuroimage.2006.09.015

Peirce, C. S. (1958). *Selected Writings: Values in a Universe of Chance*, ed. P. P. Weiner (NewYork, NY: Dover Press), 192.

Phelps, M. E., Huang, S. C., Hoffman, E. J., Selin, C., Sokoloff, L., and Kuhl, D. E. (1979). Tomographicmeasurement of local cerebral glucose metabolic rate in humans with (F-18)2-fluoro-2-deoxy-D-glucose: validation of method. *Ann. Neurol.* 6, 371–388. doi: 10.1002/ana.410060502

Pinker, S. (1997). *How the Mind Works*. NewYork, NY: W W Norton & Co.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends Cogn. Sci.* 10, 59–63.

Posner, M. I., and Raichle, M. E. (1994). *Images of Mind*. New York, NY: Freeman.

Posner, M. I., and Raichle, M. E. (1998). The neuroimaging of human brain function. *Proc. Natl. Acad. Sci. U.S.A.* 95, 763–764. doi: 10.1073/pnas.95.3.763

Price, C. J., and Friston, K. J. (2007). Functional ontologies for cognition: the systematic definition of structure and function. *Cogn. Neuropsychol.* 22, 262–274. doi: 10.1080/02643290442000095

Puce, A., Allison, T., Asgari, M., Gore, J. C., and McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings and textures: a functional magnetic resonance imaging study. *J. Neurosci.* 16, 5205–5215. doi: 10.1523/JNEUROSCI.16-16-05205.1996

Raichle, M. E. (1998). Behind the scenes of functional brain imaging: a historical and physiological perspective. *Proc. Natl. Acad. Sci. U.S.A.* 95, 765–772. doi: 10.1073/pnas.95.3.765

Raichle, M. E. (2010). *The Brain's Dark Energy*. Basingstoke: Scientific American.

Raichle, M. E. (2015). The restless brain: how intrinsic activity organizes brain function. *Philos. Trans. R. Soc. B.* 370:20140172. doi: 10.1098/rstb.2014.0172

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 676–682. doi: 10.1073/pnas.98.2.676

Reynolds, J. H., and Desimone, R. (2003). Interacting roles of attention and visual salience in V4. *Neuron* 37, 853–863. doi: 10.1016/S0896-6273(03)00097-7

Riedl, V., Bienkowska, K., Strobel, C., Tahmasian, M., Grimmer, T., Förster, S., et al. (2014). Local activity determines functional connectivity in the resting human brain: a simultaneous fdg-pet-fMRI study. *J. Neurosci.* 34, 6260–6266. doi: 10.1523/JNEUROSCI.0492-14.2014

Roth, J. K., and Courtney, S. M. (2007). Neural systems for updating object working memory from different sources:sensory stimulation or long-term memory. *Neuroimage* 38, 617–630. doi: 10.1016/j.neuroimage.2007.06.037

Rothman, D. L., De Feyter, H. M., de Graaf, R. A., Mason, G. F., and Behar, K. L. (2011). 13C MRS studies of neuroenergetics and neurotransmitter cycling in humans. *NMR Biomed.* 24, 43–957. doi: 10.1002/nbm.1772

Satel, S., and Lilienthal, S. O. (2013). *Brainwashed:The Seductive Appeal of Mindless Neuroscience*. New York, NY: Basic Books.

Schlunzen, L., Juul, N., Hansen, K. V., Gjedde, A., and Cold, G. E. (2010). Regional cerebral glucose metabolism during sevoflurane anaesthesia in healthy subjects studied with positron emission tomography. *Acta Anaesthesiol. Scand.* 54, 603–609. doi: 10.1111/j.1399-6576.2010.02208.x

Shulman, R. G. (1996). Interview with Robert G. Shulman. *J. Cogn. Neurosci.* 8, 474–480. doi: 10.1162/jocn.1996.8.5.474

Shulman, R. G. (2013). *Brain Imaging-What it Can and Cannot Tell us About Consciousness*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199838721.001.0001

Shulman, R. G., Hyder, F., and Rothman, D. L. (2009). Baseline brain energy supports the state of consciousness. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11096–11101. doi: 10.1073/pnas.0903941106

Shulman, R. G., Hyder, F., and Rothman, D. L. (2014). Neuroenergetic basis of functional neuroimaging. *J. Cereb. Blood Flow Metab.* 34, 1721–1735. doi: 10.1038/jcbfm.2014.145

Shulman, R. G., and Rothman, D. L. (1998). Interpreting functional imaging studies in terms of neurotransmitter cycling. *Proc. Natl. Acad. Sci. U.S.A.* 95, 11993–11998. doi: 10.1073/pnas.95.20.11993

Shulman, R. G., Rothman, D. L., and Hyder, F. (2007). A BOLD search for baseline. *Neuroimage* 36, 277–281. doi: 10.1016/j.neuroimage.2006.11.035

Sibson, N. R., Dhankhar, A., Mason, G. F., Rothman, D. L., Behar, K. L., and Shulman, R. G. (1998). Stoichiometric coupling of brain glucose metabolism and glutamatergic neuronal activity. *Proc. Natl. Acad. Sci. U.S.A.* 95, 316–321. doi: 10.1073/pnas.95.1.316

Smith, S. M., Vidauuree, D., Beckmann, C. F., Glasser, M. F., Jenkinson, M., Miller, K. L., et al. (2013). Functional connectomics from resting-state fMRI. *Trends Cogn. Sci.* 17, 688–682. doi: 10.1016/j.tics.2013.09.016

Stark, C. E., and Squire, L. R. (2001). When zero is not zero: the problem of ambiguous baseline conditions in fMRI. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12760–12766. doi: 10.1073/pnas.221462998

Stender, J., Kupers, R., Rodell, A., Thibaut, A., Chatelle, C., Bruno, M. A., et al. (2015). Quantitative rates of brain glucose metabolism distinguish minimally conscious from vegetative state patients. *J. Cereb. Blood Flow Metab.* 35, 58–65. doi: 10.1038/jcbfm.2014.169

Stender, J., Mortensen, K. N., Thibaut, A., Sarkner, S., Gjedde, A., and Kupers, R. (2016). The minimal energetic requirement of sustained awareness after brain injury. *Curr. Biol.* 26, 1494–1499. doi: 10.1016/j.cub.2016.04.024

Suhler, C. L., and Churchland, P. (2011). Can innate, modular "foundations" explain morality. Challenges for Haidt's moral foundations theory. *J. Cogn. Neurosci.* 23, 2103–2116. doi: 10.1162/jocn.2011.21637

Tank, D. W., Menon, R., Ellerman, J. M., Kim, S. J., Merkle, H., and Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation:functional brain mapping with magnetic resonance imaging. *Proc. Nat. Acad. Sci. U.S.A.* 89, 5951–5955. doi: 10.1073/pnas.89.13.5951

Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated informationtheroy:from consciousness to its physical substrate. *Nat. Rev. Neurosci.* 17, 450–461. doi: 10.1038/nrn.2016.44

Tressoldi, P. E., Sella, F., Coltheart, M., and Umilta, C. (2012). Using functional neuroimaging to test theories of cognition: a selective survey of studies from 2007 to 2011 as a contribution to the Decade of the Mind Initiative. *Cortex* 48, 1247–1250. doi: 10.1016/j.cortex.2012.05.024

Tsao, D. Y., Freiwald, W. A., Tootell, R. B., and Livingstone, M. S. (2006). A cortical region consisting entirely of face selective cells. *Science* 311, 670–674. doi: 10.1126/science.1119983

Uludag, K., Dubowitz, D. J., Yoder, E. J., Restom, K., Liu, T. T., and Buxton, R. B. (2004). Coupling of cerebral blood flow and oxygen consumption during physiological activation and deactivation measured with fMRI. *Neuroimage.* 23, 148–155. doi: 10.1016/j.neuroimage.2004.05.013

Uttal, W. (2001). *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain*. Cambridge, MA: MIT Press.

van der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., Raijmakers, M. E., et al. (2006). A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychol. Rev.* 113, 842–861. doi: 10.1037/0033-295X.113.4.842

van Eijsden, P., Hyder, F., Rothman, D. L., and Shulman, R. G. (2009). Neurophysiology of functionalimaging. *Neuroimage* 45, 1047–1054. doi: 10.1016/j.neuroimage.2008.08.026

Wager, T. D., and Smith, E. E. (2003). Neuroimaging studies of working memory: a meta-analysis. *Cogn. Affect. Behav. Neurosci.* 3, 255–274. doi: 10.3758/CABN.3.4.255

Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cereb. Cortex* 15, 1234–1242. doi: 10.1093/cercor/bhi006

Yourganov, G., Schmah, T., Churchill, N. W., Berman, M. G., Grady, C. L., and Strother, S. C. (2014). Pattern classification of fMRI data: applications for anlaysis of spatially distributed cortical networks. *Neuroimage* 96, 117–132. doi: 10.1016/j.neuroimage.2014.03.074

Yu, Y., Herman, P., Rothman, D. L., Agarwal, D., and Hyder, F. (2017). Evaluating the gray ad nd white matter energy budgets of human brain function. *J. Cereb. Blood Flow Metab.* 38, 1339–1353. doi: 10.1177/0271678X17708691

Zeki, S. (1993). *A Vision of the Brain*. Oxford: Blackwell.

Zhu, X. H., Zhang, N., Zhang, Y., Ugurbil, K., and Chen, W. (2009). New insights into central roles of cerebral oxygen metabolism in the resting and stimulus-evoked brain. *J. Cereb. Blood Flow Metab.* 29, 10–18. doi: 10.1038/jcbfm.2008.97

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

The reviewer AG declared a past co-authorship with the authors to the handling Editor.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 12 March 2019
doi: 10.3389/fnhum.2019.00068

Probing for Intentions: Why Clocks Do Not Provide the Only Measurement of Time

Ceci Verbaarschot, Pim Haselager and Jason Farquhar*

Centre for Cognition, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands

Having an intention to act is commonly operationalized as the moment at which awareness of an urge or decision to act arises. Measuring this moment has been challenging due to the dependence on first-person reports of subjective experience rather than objective behavioral or neural measurements. Commonly, this challenge is met using (variants of) Libet's clock method. In 2008, Matsuhashi and Hallett published a novel probing strategy as an alternative to the clock method. We believe their probe method could provide a valuable addition to the clock method because: it measures the timing of an intention in real-time, it can be combined with additional (tactile, visual or auditory) stimuli to create a more ecologically valid experimental context, and it allows the measurement of the point of no return. Yet to this date, the probe method has not been applied widely - possibly due to concerns about the effects that the probes might have on the intention and/or action preparation processes. To address these concerns, a 2 × 2 within-subject design is tested. In this design, two variables are manipulated: (1) the requirement of an introspection report and (2) the presence of an auditory probe. Three observables are measured that provide information about the timing of an intention to act: (1) awareness reports of the subjective experience of having an intention, (2) neural preparatory activity for action, and (3) behavioral data of the performed actions. The presence of probes was found to speed up mean action times by roughly 300 ms, but did not alter the neural preparation for action. The requirement of an introspection report did influence brain signals: reducing the amplitude of the readiness potential and increasing the desynchronization in the alpha and beta bands over the motor cortex prior to action onset. By discussing the strengths and weaknesses of the probe method compared to the clock method, we hope to demonstrate its added value and promote its use in future research.

Keywords: action, awareness, EEG, ERD, intention, movement, probe, RP

OPEN ACCESS

Edited by:
Joshua Oon Soo Goh, National Taiwan University, Taiwan

Reviewed by:
Chun Siong Soon, Duke NUS Medical School, Singapore
Rolf Verleger, Universität zu Lübeck, Germany

****Correspondence:***
Ceci Verbaarschot
c.verbaarschot@donders.ru.nl

Received: *26 July 2018*
Accepted: *11 February 2019*
Published: *12 March 2019*

Citation:
Verbaarschot C, Haselager P and Farquhar J (2019) Probing for Intentions: Why Clocks Do Not Provide the Only Measurement of Time. Front. Hum. Neurosci. 13:68. doi: 10.3389/fnhum.2019.00068

1. INTRODUCTION

Having an intention to act is commonly operationalized as the moment at which awareness of an urge or decision to act arises (Libet et al., 1983; Lau et al., 2007; Soon et al., 2008; Fried et al., 2011; Tabu et al., 2015; Alexander et al., 2016). Measuring this moment has been challenging due to the dependence on first-person reports of subjective experience rather than objective behavioral or neural measurements (Wolpe and Rowe, 2014; Haggard, 2019). A popular method

to measure the timing of an intention to act is the clock method of Libet et al. (1983). This method instructs participants to look at a clock and remember its configuration as soon as they experience an intention to act. This configuration is to be remembered and reported after the action has been performed. Variants of this paradigm use a stream of letters (Soon et al., 2008, 2013; Bode et al., 2011) instead of a traditional clock. Although this method is widely applied (Dominik et al., 2018; Saigle et al., 2018), it has been criticized repeatedly (Haggard, 2008; Nachev and Hacker, 2014; Navon, 2014; Wolpe and Rowe, 2014). Major critiques concern the requirement of constant introspection, the *post-hoc* nature of the intention reports and the ecological validity of the experimental task.

In 2008, Matsuhashi and Hallett came up with an alternative to the clock method. They invented a novel probing strategy to measure the experienced timing of an intention to act. Their strategy uses auditory probes that are presented to a participant at random points in time. These probes trigger a report on the awareness of an intention to act through a behavioral response. When a probe is presented and the participant is experiencing an intention, they need to refrain from acting (i.e., *veto*) and wait. Alternatively, when a probe is presented when they are not intending to act, they should simply *ignore* the probe and continue their self-paced actions. By comparing the timing of probes and consequent actions, one can determine during which time period the participant was aware of their intention to act.

We believe the probe method of Matsuhashi and Hallett (2008) provides a valuable addition to the popular clock method of Libet et al. (1983) for several reasons. First of all, the probe method does not require constant introspection: participants need to perform introspection only for a brief moment in response to a probe. Secondly, the probes are presented during action preparation, measuring the timing of an intention to act in real-time rather than *post-hoc* after action performance. Thirdly, the probe method can easily be used in combination with other visual or tactile stimuli. This provides the opportunity to use this method within a more complex environment and study a more ecologically valid experimental task. This way, the probe method can broaden our methodological repertoire so we can study intentional actions under various circumstances. Fourth, in addition to the timing of an intention to act, the probe method measures the *point of no return* (Matsuhashi and Hallett, 2008). This point of no return indicates until what time one is still able to veto an intended act.

Although the probe method of Matsuhashi and Hallett provides a valuable addition to the conventional clock method, it has not been applied widely. To the best of our knowledge, only one investigation (by us) has used this method since (Verbaarschot et al., 2016). The vast majority of researchers use the clock method to investigate the timing of an intention: Alexander et al. (2016), Banks and Isham (2009), Bode et al. (2011), Fried et al. (2011), Douglas et al. (2015), Haggard and Eimer (1999), Haggard et al. (2002), Keller and Heckhausen (1990), Miller et al. (2011), Jo et al. (2014), Lau et al. (2007), Rigoni et al. (2011), Schlegel et al. (2013), Schneider et al. (2013), Sirigu et al. (2004), Soon et al. (2008), Soon et al. (2013), Tabu et al. (2015), Wohlschläger et al. (2003), etc. Perhaps this is due to the fact that, while the clock method is not without problems, the probe method has some concerns of its own. These concerns mainly involve the effect that the probes might have on the experienced awareness of an intention to act, the timing of the performed actions and the underlying neural activity (as described in detail below). In order to address these concerns, a 2x2 within-subject design is tested. In this design, two variables are manipulated: (1) the requirement of an introspection report and (2) the presence of an auditory probe.

Three observables are measured that provide information about the timing of an intention to act: (1) awareness reports of the subjective experience of having an intention, (2) neural preparatory activity for action, and (3) behavioral data of the performed actions. The measured intention reports can consist of a *post-hoc* report on the vividness of an experienced intention, analog to the required constant introspection of the clock method (Libet et al., 1983). Alternatively, it can consist of the response to an auditory probe (i.e., ignore the probe or veto the action), as used in the probe method of Matsuhashi and Hallett (2008). The neural preparatory activity for action is recorded using an electro-encephalogram (EEG). Both the readiness potential (RP) and event-related desynchronization in the alpha (8–12 Hz) and beta (13–30 Hz) bands over the motor cortex prior to action are investigated. Both signatures have been reported to correlate with voluntary movement in previous research (Kornhuber and Deecke, 1965; Pfurtscheller and Aranibar, 1979; Libet et al., 1983; Doyle et al., 2005; Shibasaki and Hallett, 2006; Bai et al., 2011; Lew et al., 2012; Khalighinejad et al., 2018). The performed actions are measured through the timing of button presses and an electro-myogram (EMG) of the relevant arm-muscles.

Post-hoc, we investigated the brain activity prior to ignored and vetoed probes. When a probe is ignored, it means that the participant did not experience an intention to act at probe onset. We know that the RP has its onset up to 2s prior to action, whereas the awareness of an intention is reported up to 1.5s prior to action using the probe method (Matsuhashi and Hallett, 2008). This means that if a participant ignores a probe, one would expect no or a very weak RP prior to probe onset. However, when a probe is vetoed and the participant was experiencing an intention to act, one would expect to see an RP prior to probe onset. In this case the RP is time-locked to probe onset, therefore we expect it to be less pronounced (i.e., smaller amplitude) than when it would be time-locked to action onset. If we find an RP prior to vetoed probes and not prior to ignored probes, this would provide further credence to the probe method as an accurate tool for measuring the timing of an intention to act.

Before going into the details of our experiment, the next section (2) will provide additional background on the clock and probe methods. Section 3 describes our experiment which quantifies the concerns about the probe method by assessing the individual effects of the manipulated variables on each of the observables. Section 4 will compare the strengths and weaknesses of the probe method to those of the clock method. By addressing concerns and explicating the added value of the probe method, we hope to promote its use in future research.

2. COMPARING PROBES AND CLOCKS

Participants in studies that use the clock or probe method usually perform a similar motor task: a spontaneous action (e.g., a button press or brisk flexion of the hand) that is made by the participant whenever they experience an intention to do so. The difference between these methods is in the way they collect a report on the timing of an intention to act: see **Figure 1**. The clock method instructs participants to remember and report the configuration of a clock at the time of their experienced intention. The probe method uses auditory beeps to probe the participant at different moments in time for their awareness of an intention.

The probe and clock methods each have their strengths and weaknesses. First of all, the clock method requires participants to remember and report the onset of their intention after action performance Libet et al. (1983). This *post-hoc* method of gathering first-person reports seems prone to inaccuracies (Wolpe and Rowe, 2014). Moreover, the reported intention timings seem to be heavily influenced by the perceived action onset and/or the consequences of acting (Banks and Isham, 2009). The probe method measures the awareness of an intention on the spot. Participants need to respond immediately to a probe and have no need to retain the exact onset of their intention to act. Furthermore, since the awareness of an intention is measured prior to action performance, its timing cannot be influenced by the act itself or any of its potential consequences.

Second, the clock method requires continuous introspection: participants need to tune into their conscious experience to detect the slightest trace of an urge to act. This requirement seems to have an effect on the neural signatures that can be observed at that time (Lau et al., 2007). The probe method requires sporadic introspection: participants need to perform introspection for a brief moment in time in response to a probe. This happens once during a trial at most.

Third, the clock method requires an explicit intention report from a participant: participants are instructed to track the onset of their intention to act and remember and report its timing. Explicit awareness of an intention to perform a spontaneous (motor) action is not something we usually exhibit in our daily life. Requiring this awareness seems quite artificial. In contrast, the probe method uses the behavioral response to a probe to infer the time course of an intention to act. This implicit intention report softens the constraints on the level of awareness that is needed to perform the task. This situation seems similar to one in everyday life where we can explain our intentions when asked by someone.

"When you experience an intention to move, remember the configuration of the clock and report it"	"When you hear a beep and at that time you (a) do not feel an intention to act: ignore it (b) do feel an intention to act: veto"
• post-hoc intention report • requires continuous introspection • explicit intention report • requires constant stimulus monitoring • ecological validity issue: cannot be combined with other visual stimuli • in line with a discrete interpretation of an intention • does not measure the point of no return • single trial measurement of the timing of an intention to act	• real-time intention report • requires sporadic introspection • implicit intention report • no need for constant stimulus monitoring • potential for ecological validity: can be combined with additional stimuli • in line with a dynamic interpretation of an intention • measures the point of no return • multiple trials are needed to infer the average timing of an intention to act

FIGURE 1 | Comparison of the clock and probe methods.

Fourth, the required type of intended action that is studied by the clock and probe methods does not seem ecologically valid. Most investigations involve simple spontaneous actions like a self-paced hand movement (Libet et al., 1983; Sirigu et al., 2004; Matsuhashi and Hallett, 2008; Schlegel et al., 2013; Douglas et al., 2015) or a decision to add or subtract a number (Soon et al., 2013; Wisniewski et al., 2016). An alternative (more ecologically valid) task is difficult to find for the clock method. This is mostly due to the fact that it is difficult to combine the clock method with additional stimuli because focus and concentration is needed to observe a visually presented clock and remember the time of the experienced intention to act. In contrast, the probe method can easily be used in combination with other visual or tactile stimuli. This provides the opportunity to investigate the timing of an intention to act while performing actions that could be performed in everyday life. For instance, a recent study by Khalighinejad et al. (2018) made adaptations to a conventional moving-dot task in order to measure meaningful spontaneous hand movements. The target actions consist of voluntary decisions to skip a trial. Their design could be used in combination with the probe method to gain information on the timing of the intention to skip.

Fifth, the clock method seems in line with a discrete interpretation of an intention to act, whereas the probe method seems more in line with a dynamic one. The clock method asks participants to remember and report the moment at which they are aware of an urge or decision to act. This seems to assume that an intention is a discrete mental state that "pops up" in a participants mind at a specific moment in time, or at least that the awareness of that intention occurs at a discrete time (Uithol et al., 2014). The probe method questions the participant across a range in time, allowing a variety of moments at which one is aware of an urge or decision to act. Furthermore, the average onset of an intention to act measured with the probe method seems to be much earlier (about 1.5 s prior to action performance; see Matsuhashi and Hallett, 2008; Verbaarschot et al., 2016) than when it is measured with the clock method (about 0.15 s prior to action; see Libet et al., 1983; Haggard and Eimer, 1999; Haggard et al., 2002; Sirigu et al., 2004; Banks and Isham, 2009; Bode et al., 2011; Fried et al., 2011; Soon et al., 2013; Jo et al., 2014; Douglas et al., 2015; Tabu et al., 2015; Alexander et al., 2016). As argued previously (Verbaarschot et al., 2016), these findings fit better with the interpretation of an intention as a dynamic process rather than a discrete mental state. Unlike the clock method, the probe method can measure different stages in this process. However, the success of this method does come at a cost: quite a large amount of trials (±300) are required to get a good estimate of the awareness of an intention to act for a single participant. In contrast, the clock method provides a single-trial estimate of the onset of an intention.

Sixth and last, in addition to the timing of an intention to act, the probe method measures the point of no return (Matsuhashi and Hallett, 2008). When a probe is presented close to action onset, participants can no longer veto their action. According to Matsuhashi and Hallett, this inability to refrain from acting in response to a probe occurs around 0.13 s prior to action onset. The point in time at which this happens is referred to as the point of no return. Comparing the point of no return to the average brain activity at that point in time provides the opportunity to examine the stages of neural preparation for action after which action execution becomes irreversible. Schultze-Kraft et al. (2016), who found the point of no return at 200 ms prior to action onset, show that even after the onset of the Readiness Potential and alpha/beta ERD one can still veto an intended act. The clock method does not allow for any such analyses as it is unable to capture the point of no return.

Although the probe method seems to be a valuable addition to the clock method, it also raises some concerns. The requirement of an introspection report and the presence of a probe could potentially disrupt the "natural" process of intending to act in unknown ways (see **Figure 2A**). We have identified six possible scenarios that would invalidate the probe method as a tool to measure the timing of an intention to act [1]:

1. *Probes speed up brain signals*: in an experimental context in which spontaneous actions are performed in absence of a clear external stimulus on when to act, intentions to act may be based largely on spontaneous fluctuations in neural activity (Schurger et al., 2012). Whenever the neural activity crosses a certain threshold, this results in an action. Threshold crossing tends to happen at crests in the ongoing neural fluctuations. Probes may affect these fluctuations and push them over the threshold for action performance (see **Figure 2B**). This influence may happen irrespective of the current stage of development of the RP and alpha/beta ERD. If this is the case, we expect to find more variance in these neural signatures in conditions with probes (*sound + probe*) compared to those without (*control + introspect*). On the other hand, the neural signatures may be susceptible to probes only during a specific stage in their development. If this is the case, we expect to find a later onset of these signatures relative to action performance in conditions with probes compared to those without. Irrespective of these two cases: if probes speed up brain signals, actions should speed up as well. Therefore, we expect that in both these cases actions will be faster in conditions with probes than those without. It could be the case that probes affect brain signals only if participants should pay attention to them. If this is the case, we expect to find the differences described above in the *probe* condition only.
2. *Probes delay brain signals*: rather than pushing the neural fluctuations over the threshold for action performance, probes may bring neural activations back to baseline level (see **Figure 2C**). If this is the case, we expect to find more variance in the RP and alpha/beta ERD in conditions with probes (*sound + probe*) compared to those without (*control + introspect*). Moreover, we expect to find slower actions in conditions with probes compared to those without. Again, these effects may be specific to conditions in which the probe matters for the task at hand. In that case, we expect to find these differences in the *probe* condition only.

[1] We note that similar criticisms could be raised for the moving clock stimulus used in Libet-type studies.

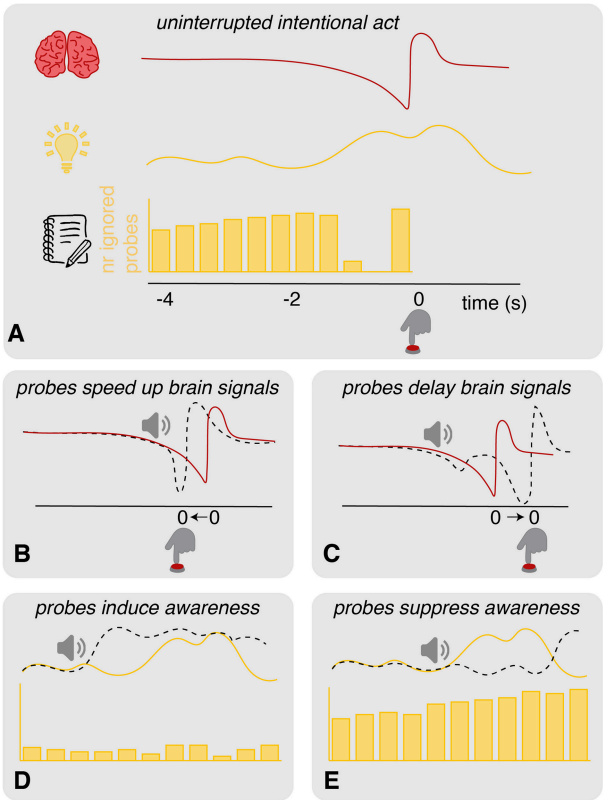


FIGURE 2 | Illustration of hypotheses. **(A)** Schematic overview of the (top: brain) brain processes (e.g., RP), (middle: light bulb) experienced intentions and (bottom: notebook) intention reports (e.g., through veto and ignore responses to probes) prior to a spontaneous act. The gap in the distribution of ignored probes is due to vetos: prior to action onset, participants are aware of their intention and perform a veto in response to any probes that are presented at this time. Shortly prior to action onset, ignored probes reappear in the distribution: at this point, probes are presented so close to action onset that participants are unable to veto. **(B)** Probes could speed up brain signals by pushing the neural activity over the threshold for action. This should speed up actions as well. **(C)** Probes could delay brain signals, bringing the neural activity back to baseline. This should delay the actions. **(D)** Probes could induce awareness of intending to act. In this case, probes would often elicit a veto response, causing a sparse distribution of ignored probes that lacks a clear gap. **(E)** Probes could also suppress the awareness of intending to act. In this case, the distribution of ignored probes closely resembles the underlying distribution of scheduled probes. A clear gap due to consistent veto responses is missing.

3. *Probes induce awareness*: the presentation of a probe may enhance the awareness of an intention to act or even cause an intention to act (see **Figure 2D**). If this is the case, almost all probes in the *probe* condition should result in a veto response. This would cause the distribution of ignored probes to look sparse.
4. *Probes suppress awareness*: probes may also suppress awareness of an intention to act (see **Figure 2E**). In this case, probes should almost never result in a veto response in the *probe* condition. The distribution of ignored probes should look very similar to the distribution of scheduled probes.
5. *Veto influences action*: participants may dislike the required veto response and may therefore attempt to act before a probe is presented. In this case, actions should be performed faster in the *probe* condition compared to the other conditions.
6. *Inaccurate intention report*: participants may simply not be able to report their intentions to act using the probes. In this case, veto's are expected to be performed randomly in response to a probe. In this case, the distribution of ignored probes of the *probe* condition should look quite similar to the distribution of scheduled probes: there is no clear time range during which vetos are consistently performed.

To quantify these six concerns, each scenario is tested using the methods described in the next section.

3. MATERIALS AND METHODS

3.1. Participants

The experiment was conducted in accordance with the ethical standards provided by the 1964 Declaration of Helsinki. The study protocol was approved by the local Ethics Committee of Faculty of Social Sciences of the Radboud University Nijmegen. A total of 21 healthy participants (15 females, mean age: 26 years old, youngest participant: 19 year old, oldest participant: 55 years old) volunteered to perform the experiment with their written informed consent. All participants were right-handed and had normal or corrected-to-normal vision. Participants received €25 or 2.5 course credits for their participation.

Five participants were excluded from the analysis. One participant reported to suffer from a brain disease that affects the amount of blood vessels present in the brain. Since it is unknown how this disease might affect their brain activity or behavior in this experiment, it was decided to exclude this participant from further analysis. Another participant reported to be nauseous during the experiment and could not sit still. Due to the large amount of resulting movement artifacts in the EEG data, this participant was excluded from further analysis. Similarly, many movement artifacts were found in the data of one other participant, leading to their exclusion from further analysis. Two participants did not follow instructions correctly, as was apparent from their answers to a post-experiment questionnaire. They were also excluded from analysis. The data of the remaining 16 participants was analyzed.

3.2. Experimental Procedure

The experiment consisted of a 2 × 2 within subject-design in which the following variables were manipulated: (1) the requirement of an introspection report and (2) the presence of an auditory probe. The main task of the participants, underlying each of the four conditions, was to press a button with the index finger of their right hand whenever they wanted to, similar to Libet et al. (1983) and Matsuhashi and Hallett (2008). Participants were instructed not to plan their actions, but press the button as soon as they felt an intention to do so. While performing these self-paced spontaneous actions, participants were instructed to look at a fixation cross that was displayed at the center of a computer screen. As long as the fixation cross was present, participants were instructed to relax, rest their arms and hands in between button presses and blink as little as possible.

When there are no reasons for deciding when to act other than a spontaneous intention to do so, it is challenging to prevent participants from acting as soon as possible (i.e., within the first 2 s of a trial). In Libet-type experiments this is done by instructing participants to wait for at least one full revolution of the clock before acting (Libet et al., 1983; Haggard and Eimer, 1999). Matsuhashi and Hallett (2008) instructed participants to perform their actions at intervals of roughly 5–10 s, without planning their actions or keeping time. When the action intervals were too long or short, participants were notified by the experimenter. Because details are missing on exactly when and how the participants were notified during the experiment of Matsuhashi and Hallett (2008), we achieved the action interval of 5-10 s using trial-by-trial color feedback: immediately at every button press, the color of the fixation cross changed for 1 s before the cross disappeared. If it turned blue, the action was made too slow; if it turned red, the action was made too fast; and if it turned green the action timing was perfect. The participant was instructed to adjust the timing of their button-presses depending on the color feedback. In this way, participants had no need to keep track of time, but could rely on the color feedback. The trained action timing provided a window of opportunity of about 5 s. During this time window, participants were free to perform a spontaneous act. This window ensured that there would be enough data for the subsequent EEG analysis and sufficient time to present a probe prior to action onset.

On top of the main task of performing self-paced button presses, each of the two independent variables were individually manipulated. This resulted in the following four experimental conditions (visualized in **Figure 3**):

1. *Control*: participants performed the main task of pressing a button at their own pace roughly every 5–10 s. An introspection report on their experienced intention was not required. This was the most basic condition as it consisted solely of the performance of spontaneous voluntary actions. With no additional stimuli or mental tasks, this condition provided pure control data for the timing of intended actions and their preceding neurological signals.
2. *Sound*: in addition to the main task, an auditory probe was presented at random times. Participants were instructed to ignore this probe completely because it has no importance to the experiment. Again, an introspection report on their experienced intention was not required. This condition allowed the investigation of any potential effects of the added auditory stimuli on the neural preparation for action, the awareness of an intention and the action itself.
3. *Introspect*: This condition did not involve any probes, but did require an introspection report. In addition to the main task, participants were instructed to focus their attention on the first moment at which they felt the urge to press the button. Immediately at every button press, the following multiple-choice question was presented: "How did you experience your intention to act?". Participants could answer this question by pressing one of three buttons corresponding to the following answer options: "vivid and conscious," "a vague feeling of wanting" or "pressed the button without thinking about it." To prevent action preparation prior to the presentation of this question, the order of these answer options was set randomly at the start of each trial. Using these instructions and questions, participants were required to maintain a constant meta-awareness of their intention to act. This requirement mimicked the level of introspection required in a Libet-type experiment (Libet et al., 1983) and allowed the investigation of any effects of the pure introspection task - without the additional visual stimuli and memory tasks required by the clock method. Color feedback on action timing was provided immediately after answering the multiple-choice question.
4. *Probe*: in addition to the main task, an auditory probe was presented at random times and an introspection report was required. When the probe is presented while (1) they had an intention to act: they should *veto* the intended act (i.e., not press the button) and wait for the fixation cross to disappear. Alternatively, when a probe is presented while (2) they did not have an intention to act: they should *ignore* the probe and press the button whenever they wanted to. These instructions were a direct replication from Matsuhashi and Hallett (2008). Whenever a trial ended without a button press, the question "did you intend to act at the time you heard the beep?" was presented. The participant could answer this multiple-choice question with either "yes" or "no." To prevent action preparation prior to the presentation of this question, the order of these answer options was determined randomly at

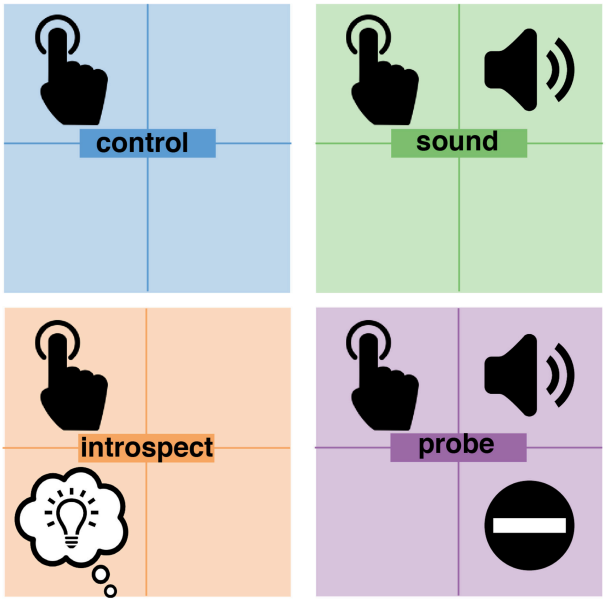


FIGURE 3 | Overview of experimental conditions. In all conditions, participants were pressing a button at their own pace. In addition, the *sound* and *probe* conditions presented an auditory probe to participants at pseudo-random moments in time. The *introspect* and *probe* conditions both required a report on the awareness of an intention to act. In the *introspect* condition, participants needed to report *post-hoc* how vividly they experienced their intention to act. In the *probe* condition, participant needed to veto their action in response to a probe when at that time they were aware of their intention to act.

the start of each trial. This question was presented in order to distinguish a veto from the absence of a button press (i.e., the trial ended before the participant experienced an intention to act). The subjective experience of an intention was reported indirectly through the behavioral response (i.e., veto or ignore) to a probe and was confirmed by answering the question at the end of a trial. This was the core condition that implemented the full probe method. Color feedback on action timing was provided immediately at each button press or, in case a trial ended without a button press, after answering the multiple-choice question.

At the end of the experiment, participants 8 to 21 completed a short questionnaire in order to gain more insight into their subjective experience during the experiment. The questionnaire consisted of the following seven questions (translated from Dutch):

1. Did you act spontaneously?
   a. Yes, I pressed the button as soon as I wanted to
   b. No, long before I pressed the button, I already decided to act
2. Was the difference between the tasks clear?
   a. Yes
   b. No
3. What did you think about the beeps?
   a. Annoying
   b. Neutral
   c. Stressful
   d. Other: ...
4. Was it difficult to determine whether you had an intention to act at the time you heard the beep?
   a. Yes
   b. No
   c. Sometimes
5. How did you decide whether you could press the button after you heard a beep?
6. Was there a clear difference between the moment at which you had an intention to press the button and the moment at which you pressed the button?
   a. Yes, the intention to press the button was clearly distinguishable from the button press itself
   b. No, the intention to press the button occurred at the same time as the button press
   c. Other: ...
7. Did the beeps influence your intention to press the button?
   a. No
   b. Yes, because: ...

3.3. Stimuli

The participant was seated on a comfortable chair in front of a table inside a quiet room. The instructions and visual stimuli were displayed using a 17 inch TFT screen with a resolution of 800 by 600 pixels and a refresh rate of 60Hz that was placed roughly at 70cm directly in front of the participant. In-ear headphones were used to present the auditory probes. A button box containing a total of four buttons was used to perform the self-paced button presses and answer the questions in the *introspect* and *probe* conditions. The experiment was run in BrainStream [2].

The auditory probe consisted of a short "beep" that was created in Matlab [3]. The probe had a frequency of 1200 Hz and duration of 0.04 s. Matsuhashi and Hallett (2008) state that "Tones were applied pseudo-randomly at intervals of 3–20 s, controlled by one of the investigators in a way that was not predictable by the subjects" (pp. 2345–2346). However, because further details on the exact timing of the probes are missing, these probes times are not replicable. For this reason we designed our own probe distribution. The timing of our probes are predetermined on the basis of 25 *control* trials that were collected during a training block at the start of the experiment. The probe onsets ranged from 0.5 s before the average action time plus and minus one standard deviation. Within this interval, auditory probes followed a truncated normal distribution such that most probes were presented before the average action time (see **Figure 4**). A minimum probe interval of 3 s before the average action time was ensured. Moreover, the probe interval was ensured to start at least 3 s after trial onset. As well as being explicable, this probe distribution was designed to optimize experimental efficiency by ensuring that approximately one third of all trials would present a probe within 3 s before movement onset (during which awareness of an intention to act is most likely to occur). Both the *sound* and *probe* conditions used the exact same probe distribution per participant. Every trial in the *sound* or *probe* conditions could contain maximally one probe. Depending on the participants action time, this probe may or may not be presented on a certain trial.

The color feedback was slightly random. The fixation cross turned red (i.e., too fast) if the button press was made within the first 5 s after trial start + a random time interval between 0 and 3s. The fixation cross turned blue (i.e., too slow) if the button press occurs more than 10 s after trial start + a random time interval between 0 to 3 s. In all other cases where a button press was made, the fixation cross turned green. This means that the boundaries between each feedback color were blurred by 0–3 s. These blurred boundaries within the color feedback were designed to encourage the element of spontaneity. They made it difficult for a participant to count time or otherwise plan their actions to perform them at a "correct" time. Instead, they needed to refrain from planning and focus on their intention to act within the learned window of opportunity.

The presentation of the fixation cross marked the start of a trial. Each trial had a maximum duration of 9 to 14 s. The exact trial duration was chosen randomly at the start of each trial. A trial ended either because the participant had pressed the button or because the maximum trial duration was reached. The

[2] see www.brainstream.nu
[3] see www.mathworks.com

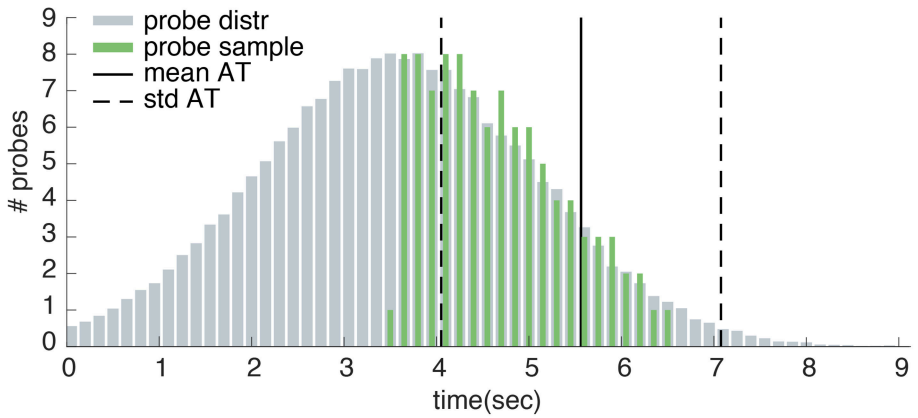


FIGURE 4 | Example probe distribution. Here, the measured action times during training have a mean of 5.563 s (mean AT) and a standard deviation of 1.513 s (std AT). Probes are sampled from a normal distribution with a mean of mean AT - std AT - 0.5 and a standard deviation equal to std AT. The sampled probe onsets follow a truncated normal distribution within the interval of 0.5s before the average action time plus and minus one standard deviation.

inter trial interval was chosen randomly between 1.5 and 3 s on each trial.

3.4. Experimental Timeline

At the start of the experiment, participants performed two training blocks. The first block consisted of 20 *control* trials and was used to train participants on the main task of pressing a button whenever they wanted to. This block was repeated until the participant performed the desired actions at roughly the right time interval. The second block consisted of 25 *control* trials and was used to collect the action times required to set the time distribution of the probes. The remainder of the experiment consisted of 4 test sequences. Each test sequence consisted of 4 blocks of 25 trials of each condition in a random order. The type of condition was displayed to the participant prior to each sequence of 25 trials of a single condition. In total, 100 trials were acquired per condition.

The experiment took 1.5 to 2 hours + 0.5 hours for setting up the EEG, which resulted in a maximum total duration of 2.5hours.

3.5. Behavioral Data

Three behavioral measurements were collected during the experiment: (1) the introspection reports, (2) the timing of the performed actions and (3) the answers to the questionnaire. In the *introspect* condition, the introspection reports consisted of the experienced vividness of the intention to act. In the *probe* condition they consisted of a behavioral response to a probe (a veto or ignore) and its confirmation at the end of a trial. The action timing was measured using the button presses and the onset of muscle activation as recorded with an EMG.

In order to quantify hypotheses 3, 4 and 6 (see Section 2), we needed to determine whether the probes in the *probe* condition lead to veto responses across a consistent time range prior to action onset. In other words: is there a gap in the distribution of ignored probes, as illustrated in **Figure 2A** and found by Matsuhashi and Hallett (2008)? To answer this question, the timing of the ignored and scheduled probes relative to action onset was analyzed. The distribution of scheduled probes refers to the average amount of probes that should by design occur prior to action onset, as described in Section 3.3. Since participants performed self-paced actions during the experiment, one could not predict their exact action onsets. Therefore, the amount of probes that were actually presented prior to action onset will always differ a bit from the scheduled ones. The distribution of ignored probes is a sample from the scheduled probe distribution that shows how many of the scheduled probes were actually presented and ignored by the participant.

To determine the distribution of scheduled probes relative to action onset, the scheduled probe onsets that precede each individual action were sampled per participant. Subsequently, the action onset was subtracted from each corresponding sampled probe onset in order to calculate the probe timing relative to action onset. A histogram with 33 time bins of 150 ms, running from 5 s prior to action until action onset, was constructed of the scheduled probe timings. In order to get an estimate of the average number of scheduled probes per participant, the histogram of scheduled probes was divided by the total number of performed actions. In addition, a histogram was created of the ignored probe times (probes that were followed by an action at a later point in time) using identical time bins. Lastly, the mean distribution of scheduled and ignored probes was calculated across all participants. A Wilcoxon signed-rank test (Wilcoxon, 1945) was used to assess whether the values of each time bin differ significantly between the ignored and scheduled probes across participants. The alpha-level for significance was Bonferroni corrected and set at $0.05/33 = 0.0015$. The consecutive time points at which the number of ignored probes were found to be significantly less than the number of scheduled probes define a time range during which participants were on average aware of their intention to act (see **Figure 6**). As a control, these steps were repeated for the *sound* condition (which should not show a gap in the distribution of ignored probes prior to action onset).

The EMG measurement served to check the accuracy of the button presses. The EMG was recorded using two electrodes, placed in a bipolar pair on the right wrist and the center of the right forearm (on the flexor pollicis longus). For analysis, the EMG data of the bipolar electrodes was subtracted and band-pass filtered between 51 and 250 Hz. Subsequently, the absolute value was taken and the data was sliced in epochs of 4 s prior to a button press until 1 s after. For each participant, the average EMG activity was calculated across all trials. From this average the median and standard deviation were calculated for each participant over all time points. An individual threshold for muscle activation was set at the median EMG activity plus 2x its standard deviation. The average onset of muscle activity was determined as the point in time at which the average EMG activity exceeded the set threshold.

The mean and standard deviation of the action times relative to trial start were calculated for each participant. To quantify hypotheses 1, 2, and 5 (see Section 2), a two-factor within-subject repeated measures ANOVA was used to assess any significant effects of the manipulated variables on the mean and standard deviation of the action times between conditions. The two factors are: (1) the requirement of an introspection report and (2) the presence of probes. By using a Bonferroni correction for these two factors and their potential interaction, the significance level was set to $0.025/3 = 0.008$ for a two-sided significance test. When a main effect of either manipulated variable was found, individual *post-hoc* paired-samples t-tests were used to assess specific differences in action mean or standard deviation between conditions. The significance level of these individual tests was set to 0.025 for a two-sided significance test.

In addition, the relation between probes and actions was investigated by looking at differences between conditions in action times relative to probe onset. In order to calculate the action times of the *control* and *introspect* conditions relative to probe onset, the same probe distribution was used as presented in the *sound* condition. These simulated probe onsets—button press times represent the absence of a connection between probes and actions in the *control* and *introspect* conditions and served as a control for the *sound* condition in which such a relation may be present. Differences in action mean or standard deviation relative to probe onset were assessed using a paired-sample t-test on all three possible combinations of the *control*, *introspect* and *sound* conditions. The significance level was set to $0.025/3 = 0.008$. The *probe* condition was left out of this analysis since it would differ by design from all other conditions due to the performed veto responses: creating a potential gap in the distribution of ignored probes, as illustrated in **Figure 2A**.

3.6. Brain Data

EEG data was collected using 64 Ag/AgCl active electrodes sampled at 512 Hz using Biosemi equipment [4]. The electrodes were placed according to the International 10/20 system. The electrodes measured all frequencies between 0 and 512 Hz. Two electrodes were placed on the left and right mastoids and four electro-oculogram (EOG) electrodes were placed in bipolar pairs above and below the left eye and on the outer sides of both eyes. The neural data was analyzed using Fieldtrip [5].

The EEG data was preprocessed using the following steps:

1. Data was sliced in epochs of 10 s before to 5 s after action onset (i.e., button press), so the data was time locked to action onset (at 0 s).
2. Data is downsampled to 256 Hz.
3. Trials in which the participant acted within 4 s after trial start were removed to ensure a decent baseline period.
4. Trials in which a probe was presented between 4.5 and 2.5 s before action onset were removed to ensure a decent baseline period.
5. Data of all conditions was concatenated per participant.
6. Data was rereferenced using a linked-mastoid reference.
7. Baseline correction was performed per trial and electrode by subtracting the average EEG signal between 3.5 and 2.5 s prior to action onset.
8. EOG artifacts were removed using a linear decorrelation of the recorded EEG and EOG (Gratton, 1998).
9. A band-pass filter between 0.2 and 47 Hz was used to filter out slow drifts and 50 Hz line noise.
10. Epochs of -5 to 3 s around action onset were retained.
11. Bad channels were removed if they differed more than 3.5 standard deviation in power from the median across all channels.
12. Bad epochs were removed if they differed more than 3.5 standard deviation in power from the median across all trials.
13. Bad channel rejection was repeated.
14. A spherical spline interpolation was used to reconstruct bad channels (Perrin et al., 1989).

To quantify hypotheses 1 and 2 (see Section 2), individual Event-Related Potentials (ERPs) were calculated per participant and per condition. To assess the main effects of the requirement of an introspection report and the presence of probes on the RP, mean ERPs were calculated across the following conditions:

1. *Control* and *sound*: providing information about all conditions without introspection reports.
2. *Introspect* and *probe*: providing information about all conditions with introspection reports.
3. *Control* and *introspect*: providing information about all conditions without probes.
4. *Sound* and *probe*: providing information about all conditions with probes.

Two within-subject cluster permutation tests with 1000 permutations were used to assess whether the last 2.5 s of data prior to action onset differed between these grouped conditions: introspection vs. no introspection and no probe vs. probe conditions (Maris and Oostenveld, 2007). After a Bonferroni correction, the significance level was set to $0.025/2 = 0.013$ for a two-sided significance test. If significant main effects were found, *post-hoc* within-subject cluster permutation tests were used to identify significant differences between individual experimental

[4] see www.biosemi.com

[5] see www.fieldtriptoolbox.org

conditions. The significance level of these individual tests was set to .025 for a two-sided significance test.

Data from the *probe* condition was used to analyze the RP prior to ignored and vetoed probes, since this was the only condition that contained both ignore and veto responses to probes. From this data, four types of trials were extracted (see **Figure 5**):

1. *Action-without-probe*: trials in which no probe occurred and an action was performed at least 4 s after trial start (to ensure a decent baseline).
2. *Ignore*: trials in which a probe occurred at least 4 s after trial start and was followed by an action more than 0.5 s later (which was determined as the point of no return, as shown in Section 4.1.2).
3. *Veto*: trials in which a probe occurred at least 4 s after trial start and was followed by a veto (i.e., no action). The veto response was confirmed by the answer to the veto question at the end of the trial.
4. *Incorrect action*: trials in which a probe occurred at least 4 s after trial start and was followed by an action within 0.5 s.

According to Matsuhashi and Hallett (2008), vetos should consistently be performed across a specific time range prior to action onset. Moreover, this time range should coincide with the build-up of the RP (Verbaarschot et al., 2016). If this is the case, we expect to find a weak RP signature prior to vetoed probes and no RP prior to ignored probes (see **Figure 6**). To test this, action-without-probe trials were time-locked to the performed button press, whereas the ignore, veto and incorrect action trials were time-locked to probe onset. To extract the general signature of the RP (i.e., a negative potential relative to baseline), the mean activities during the last 1.5 s to 0.5 s and 0.5 s to 0 s prior to action or probe onset were calculated per participant at electrode Cz. These two time intervals were used to assess whether there are differences in the early and/or late phase of the RP between these different trial types (Shibasaki and Hallett, 2006). For the two time intervals, two-sided dependent samples t-tests were used to test whether the activity differs significantly between action-without-probe and incorrect action trials and action-without-probe and veto trials (Bonferroni corrected at $0.025/8 = 0.003$). A further one-sided dependent samples *t*-tests were used to test for both time intervals whether the activity was more negative in veto and action-without-probe trials relative to ignore trials (Bonferroni corrected at $0.05/8 = 0.006$).

To quantify hypotheses 1 and 2 (see Section 2), individual alpha/beta ERDs were assessed. For this reason, the data of all conditions was concatenated and a spectogram was calculated between −5 and 3 s around action onset. Frequencies of interest were defined from 5 to 30 Hz using 2 Hz bins. A flexible Hanning window was used such that it included at least 7 cycles of each frequency of interest. The data was baselined using a relative baseline, resulting in the relative signal change compared to baseline (where a value of 1 means no change). The baseline activity was defined per electrode, frequency and trial as the median power between 3.5 and 2.5 s prior to action. Subsequently, the data was separated into the different conditions. The ERD was calculated per participant by taking the median power across trials for each electrode, frequency and trial. Mean ERDs were calculated per participant for no introspection (*control* and *sound*) vs. introspection conditions (*introspect* and *probe*), and no probe (*control* and *introspect*) vs. probe conditions (i.e., *sound* and *probe*). Within-subject cluster permutation tests (Maris and Oostenveld, 2007) with 1000 permutations were used to assess whether the last 2.5 s of data prior to action onset differed between the no introspection vs. introspection and no probe vs. probe conditions. After a Bonferroni correction, the significance level was set to $0.025/2 = 0.013$. When a significant main effect was found, *post-hoc* within-subject cluster permutation tests were used to identify significant differences between individual experimental conditions. The significance level of these individual tests was set to .025 for a two-sided significance test.

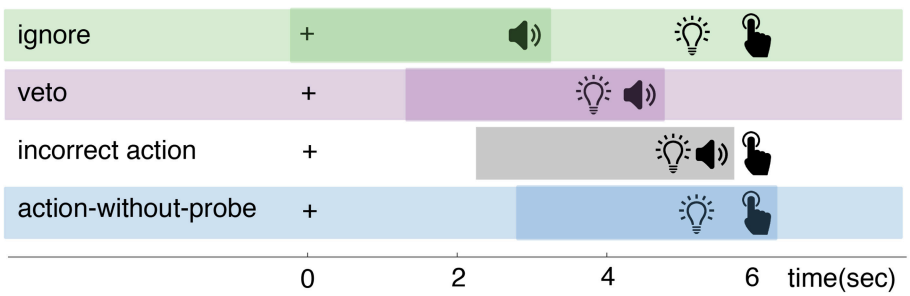


FIGURE 5 | Schematic timeline of events that could happen within a trial of the *probe* condition. The "+" indicates trial start. Each *probe* trial could develop in one of four ways. (1) Ignore: a probe is presented when the participant is not intending to act. The participant ignores the probe and presses the button at a later point in time when he/she feels the intention to do so. (2) Veto: a probe is presented when the participant experiences an intention to act. The participant vetos his/her action and waits for the trial to end. (3) Incorrect action: a probe is presented when the participant experiences an intention to act. The probe happens so close to action onset, that the participant is unable to veto his/her action and presses the button anyway. (4) Action-without-probe: the participant presses a button when he/she feels the intention to do so. They are uninterrupted by a probe. The darker colored portions of each row indicate the portion of the trial that is used for EEG analysis. This portion is either time-locked to probe (ignore, veto and incorrect action trials) or action onset (action-without-probe trials).

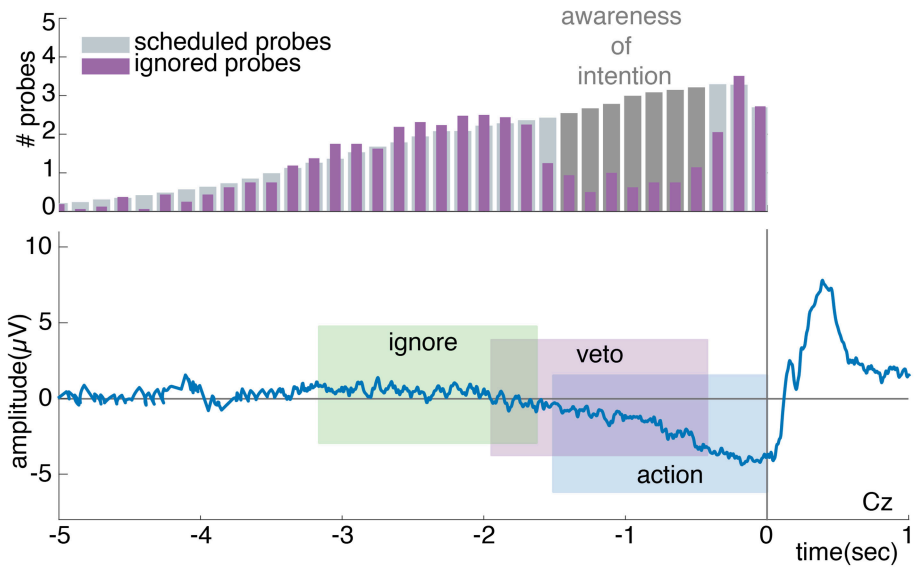


FIGURE 6 | The top plot shows a schematic grand average of the distribution of scheduled and ignored probes relative to action onset. The difference between the scheduled and ignored probe distribution, as highlighted in dark gray, shows the time points at which participants are expected to perform a veto in response to a probe. The gap provides an estimated time period during which participants are expected to be aware of their intention to act. The bottom plot shows a schematic grand average of the expected RP. When looking at the average brain activity prior to an action, a full RP is expected to be present. When looking at the average brain activity prior to an ignored probe, no RP is expected. The analysis window prior to a vetoed probe is expected to sample the early phase of the RP, which given the shape of the RP will look like a weaker version of the full RP.

4. RESULTS

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

4.1. Behavioral Data

4.1.1. Questionnaire

A total of 14 participants completed the questionnaire at the end of the experiment (see Section 3.2). Due to the exclusion of some participants (see Section 3.1), the answers of a total of 12 participants were analyzed. All 12 participants indicated that the differences between the *control*, *sound*, *introspect* and *probe* conditions were clear. Eleven participants indicated that they acted spontaneously during the entire experiment, i.e., they pressed the button as soon as they wanted to. In contrast, 1 participant indicated that he/she did not act spontaneously, i.e., he/she already decided to act long before they pressed the button.

Six participants always experienced a clear difference in timing between an intention and action, whereas 1 participant only experienced this sometimes and 1 other participant only had this experience when he/she perceived the intention consciously and vividly. One participant experienced the intention as always occurring prior to the action. Three participants did not experience any difference in timing between the intention and action and perceived them as occurring at the same time.

Concerning the *sound* and *probe* conditions, 7 participants reported the probes as neutral: they did not experience any positive or negative effects caused by the probes. However, 5 participants experienced some negative effects from the probes as they reported them to be "annoying" (2 participants), "stressful" (2 participants), or as "disturbing their relaxed state" (1 participant). Moreover, 9 participants believe that the probes did influence their course of action, whereas 3 participants did not experience any influence of the probes.

During the *probe* condition, 8 participants found it *sometimes* difficult to judge whether or not they were experiencing an intention to act when a probe was presented. One participant always experienced this intention assessment as difficult and 3 participants had no trouble with it at all. Eight participants followed instructions correctly and vetoed their intended movement when they experienced an intention to act upon probe presentation, whereas another 2 participants did not use any particular strategy to decide whether or not they could press the button after a probe was presented. Two participants determined whether or not they should veto their act based on their expectation of a probe. Whenever they were expecting a probe and a probe was presented, they would veto their subsequent act. In their case, the performed veto's did not relate to their intention to act but to their expectation of a probe. Because these participants were effectively not following instructions correctly, they were excluded from further analysis (as indicated in Section 2.1).

In summary, 92% of the participants who completed the questionnaire acted spontaneously throughout the experiment. Although 75% of the participants at least sometimes perceived their intentions and actions as two different events in time, 75% of participants also found it difficult to assess upon probe presentation whether or not they were intending to act. Lastly,

the probes seem to be experienced in a negative way by at least 42% of participants.

4.1.2. Intention Reports

Each participant completed 100 trials of the *introspect* condition. Twenty-three (=1%) trials across all participants ended without a button press. Participants reported having experienced their intention as "vivid and conscious" in 45% (±2%) of all trials containing an action. In 34% (±1%) the intention was experienced as a "vague feeling of wanting." In the remaining 21% (±1%), participants reported to have "pressed the button without thinking about it." **Figure 7A** shows the reports on the subjective experience of intending to act for each participant and across all participants.

Similar to the *introspect* condition, 100 trials were collected per participant in the *probe* condition. In 63% (±1%) of all trials across participants, a probe was presented. Seventy-eight% (±16%) of the presented probes across participants were followed by an ignore response, whereas 20% (±14%) was followed by a veto response. Three % (±5%) of the presented probes were followed neither by an ignore or a veto response; the trial simply ended before an action was made. **Figure 7B** shows the number of ignore and veto responses for each participant and the percentage of these responses across all participants.

Figures 8A,B show the number of observed ignored probes and scheduled probes relative to action onset across all participants for the *probe* and *sound* conditions. As noted in Section 3.5, the distribution of observed ignored probes will always differ a bit from the distribution of scheduled probes since the scheduled probes are an approximation of the observed probes: i.e., the predicted amount of probes that on average should occur prior to action onset. Furthermore, the distribution of ignored probes includes the amount of probes that are presented and followed by an action only, whereas the distribution of scheduled probes also includes probes that are not followed by an action (i.e., a veto).

Since our probes do not follow a uniform but a truncated normal distribution, our plot looks slightly different from those of Matsuhashi and Hallett (2008). Similar to Matsuhashi and Hallett (2008), we observe a significant decrease in the fraction of ignored probes between 1.4 and 0.65 s prior to action ($p < 0.0015$) for the *probe* condition. During this time period, participants mostly performed a veto in response to a presented probe. Shortly prior to action, from 0.5 s to action onset, the fraction of ignored probes increases again. This increase could be due to the point of no return as defined by Matsuhashi and Hallett (2008). At the point of no return, a probe was presented in such close proximity to action onset that participants are not able to cancel their action anymore to perform a veto. As shown in **Figure 8C**, the distribution of presented probes of the *sound* condition also shows a slight decrease in the amount of probes between 1 and 0.5s prior to action. However, this deviation between the scheduled and presented probes was not found to be significant.

4.1.3. Action Distribution

Across all participants, the mean onset of muscle activation was found at 81 ms (±78*ms*) prior to a button press. Since

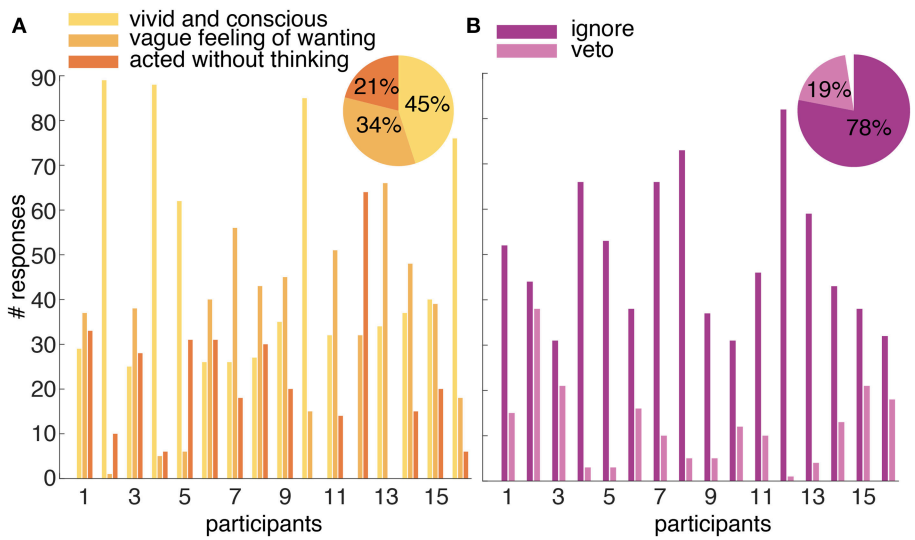


FIGURE 7 | (A) Reported experience of an intention to act in the *introspect* condition. Participants could report their intention to act after every trial as "vivid and conscious," "a vague feeling of wanting," or "pressed the button without thinking about it." The number of times each possible answer was selected is shown per participant. The percentage of each selected answer across all participants is shown in the pie chart in the top right corner. **(B)** Overview of the responses to a probe in the *probe* condition. The number of times a presented probe is followed by an ignore or a veto response is shown for each participant. The percentage of presented probes that were followed by an ignore or veto response across all participants is shown in the pie chart in the top right corner. The gap in the pie chart shows the percentage of presented probes that were ignored but not followed by an action (i.e., the participant did not intend to act during the trial).

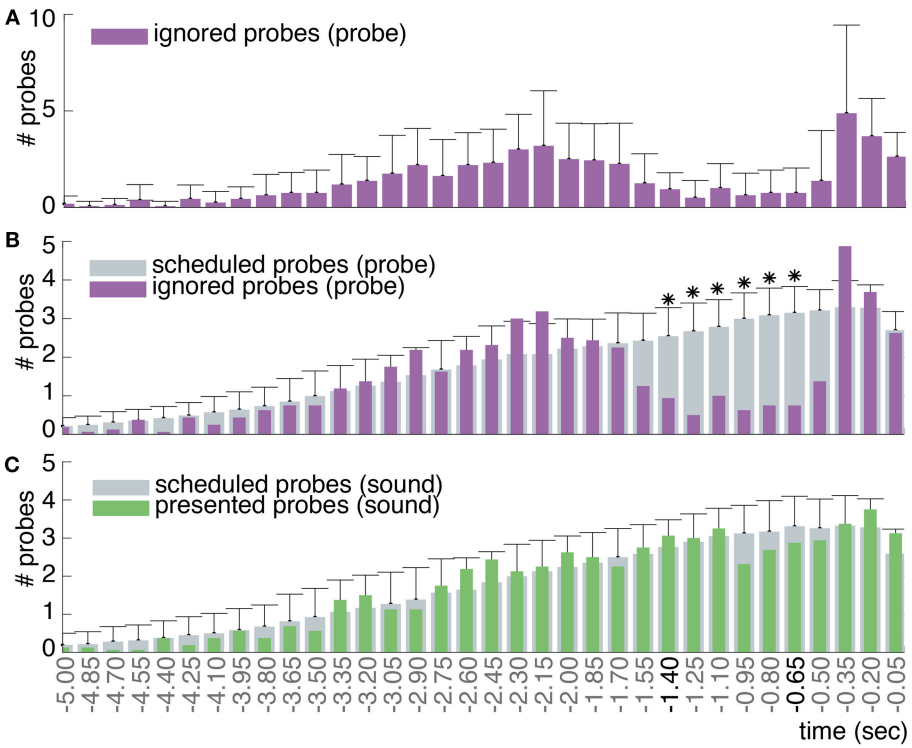


FIGURE 8 | (A) The mean and standard deviation of the ignored probes across all participants in the *probe* condition. **(B)** In gray the mean number of probes that were scheduled to be presented prior to action across all participants in the *probe* condition (standard deviations across participants are provided on top of the histogram). In purple the mean number of probes that were ignored and followed by an action (at time 0) across all participants. Around −1.4 s prior to action, participants start to veto their action in response to a probe: this is apparent by the decrease in the observed fraction of probes that were followed by an action compared to the scheduled fraction of probes. Around –0.5 s prior to action, participants are no longer able to veto their action in response to a probe due to their close proximity in time (i.e., point of no return), hence the increase in ignored probes. Note: the purple distribution of ignored probes in **(B)** is identical to that in **(A)**, but with a different scaling factor. Significant ($p < 0.0015$) differences between the amount of scheduled and ignored probes are indicated with an asterisk (*). **(C)** In gray the mean number of probes that were scheduled to be presented prior to action across all participants in the *sound* condition. In green the mean number of probes that were actually presented and followed by an action (at time 0) across all participants.

this difference is small relative to the magnitude of the expected intention reports (about 1 s) and estimation errors, the timing of the button press is used as action onset throughout the analysis.

A boxplot of the action times of the *control* (mean: 7.501 s, standard deviation: 1.549 s), *sound* (mean: 7.180 s, standard deviation: 1.447 s), *introspect* (mean: 7.720 s, standard deviation: 1.618 s) and *probe* (mean: 7.179 s, standard deviation: 1.493 s) conditions across all participants is shown in **Figure 9A**. A significant main effect of the presence of probes on the mean action time was found ($df = 15, F = 15.704, p = 0.0013$). Individual *post-hoc* tests reveal that this is due to significant differences between the *probe* and *control* ($df = 15, t = -3.100, p = 0.007$), *sound* and *introspect* ($df = 15, t = 3.823, p = 0.002$), and *probe* and *introspect* conditions ($df = 15, t = -4.199, p = .000$). No significant main effect of the requirement of an introspection report on mean action time was found ($df = 15, F = 3.291, p = 0.090$). Furthermore, no significant main effects of the requirement of an introspection report or the presence of probes was found on the standard deviation of action times.

Furthermore, the effect of the presence of a probe on action timing was explicitly investigated by comparing the action timings relative to probe onset across all conditions. A boxplot of the action timings relative to probe onset of the *control* (mean: 0.886 s, standard deviation: 1.801 s), *sound* (mean: 0.574 s, standard deviation: 1.704 s) and *introspect* (mean: 1.110 s, standard deviation: 1.840 s) conditions across all participants is shown in **Figure 9B**. Again, the mean action time of the *sound* condition was found to differ significantly from that of the *introspect* ($df = 15, t = 3.8226, p = .002^*$) condition. No significant differences in mean action time were found between the *control* and *introspect* or *control* and *sound* conditions. No significant differences in the standard deviation of action times were found between any of the conditions.

4.2. Brain Data

After preprocessing, 85 (minimum: 40, maximum: 91) trials remained for analysis of the *control* condition, 85 (minimum: 46, maximum: 92) trials for the *sound*, 87 (minimum: 48, maximum:

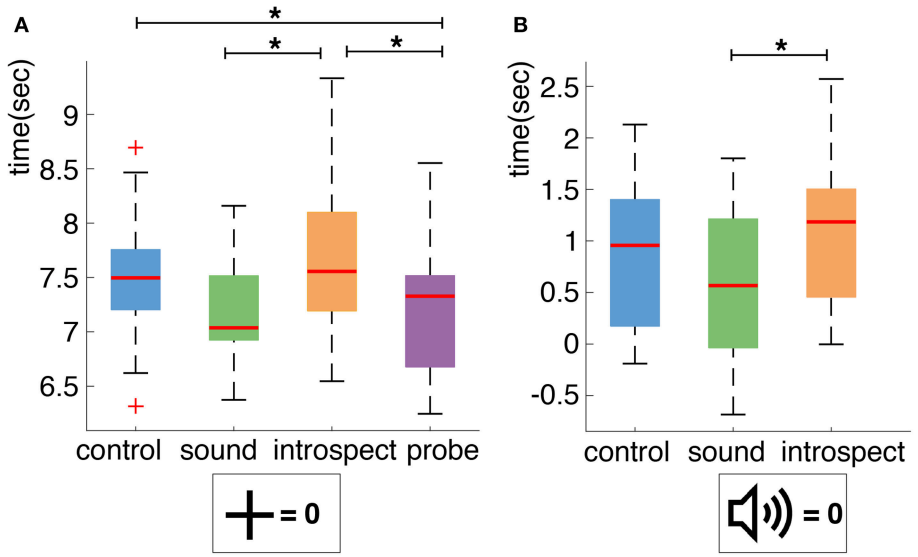


FIGURE 9 | (A) Boxplot of action times per condition across all participants. Action times are measured relative to trial start. **(B)** Boxplot of action times per condition across all participants. Action times are measured relative to probe presentation. Significant differences ($p < 0.025$ for **A** and $p < 0.008$ for **B**) between conditions are indicated with an asterisk (*).

96) trials for the *introspect* and 75 (minimum: 28, maximum: 89) trials for the *probe* condition.

4.2.1. Readiness Potential

Figure 10 shows the grand average of the RP for conditions with and without introspection reports and with and without probes. Visually, the RP seems to have its earliest onset around 2 s prior to movement. The shape and timing of the RP confirm previous research involving spontaneous voluntary right hand movements (Kornhuber and Deecke, 1965; Libet et al., 1983; Shibasaki and Hallett, 2006). Significant main effects of the requirement of an introspection report and the presence of probes are found on the last 2.5 s of the RP ($N = 16, p < 0.008$, see **Figure 10**). Whereas probes seem to cause a slight increase in RP amplitude, the requirement of an introspection report seems to cause a slight decrease in RP amplitude. The *introspect* condition seems to lie at the heart of these effects, as the RPs in this condition were found to differ significantly from those in all other conditions ($N = 16$, $p < 0.025$). **Figure 11** shows the grand average of the RP for each individual condition.

After action onset, the RP of the *introspect* condition deviates from the others (see **Figure 11**). This is due to a difference in events after action onset: immediately after action performance in the *introspect* condition, participants are prompted with a question about the vividness of their intention to act and need to respond to this question by pressing one of three buttons.

Concerning the *post-hoc* analysis of the RP prior to vetoed and ignored probes, **Figure 12** shows the grand average ERP for different trials of the probe condition. The number of ignored and vetoed probes differ greatly among participants within this condition (see Section 4.1.2). Especially the number of vetoed probes is quite low: around 20%. For this reason, we removed some participants from further analysis: only those participants who retained at least 10 trials action-without-probe, ignore and veto trials were kept for further analysis. This resulted in a total of 8 participants containing on average 39 (minimum: 30, maximum: 53) action-without-probe trials, 25 (minimum: 19, maximum: 32) ignore trials, 15 (minimum: 10, maximum: 20) veto trials and 11 (minimum: 5, maximum: 18) incorrect action trials per participant.

Figure 12A shows a clear negativity across the motor cortex for the action-without-probe and incorrect action trials. This is to be expected since both trial types include an intended action and thus should look similar to the RP in **Figure 11**. The incorrect action trials are a weaker version of the action-without-probe trials because they contain less data and are time-locked to probe onset rather than action onset. As the RP increases in amplitude up to action at 0 s, time-locking to earlier times in the RP development results in reduced ERP amplitudes, as seen here for the incorrect action trials where probes occur randomly somewhere in the last 0.5 s before action onset. The veto trials show a medium and more widespread negativity across the motor cortex, whereas this negativity seems completely absent for ignore trials. A similar trend can be observed from **Figure 12B**: a clear negativity (i.e., RP) can be observed for action-without-probe trials, whereas ignore trials remain around baseline. Veto trials (which we expect are randomly sampled from –1.4 to –0.65 s before action based on the intention-window identified in **Figure 8**) are in between ignore and action-without-probe trials, showing a medium negativity shortly prior to probe onset.

One-sided dependent samples *t*-tests showed that the action-without-probe trials are significantly higher in amplitude compared to the ignore trials during both the early (–1.5 to –0.5 s: $df = 7, t = -3.5217, p = .005$) and late (–0.5 to 0 s:

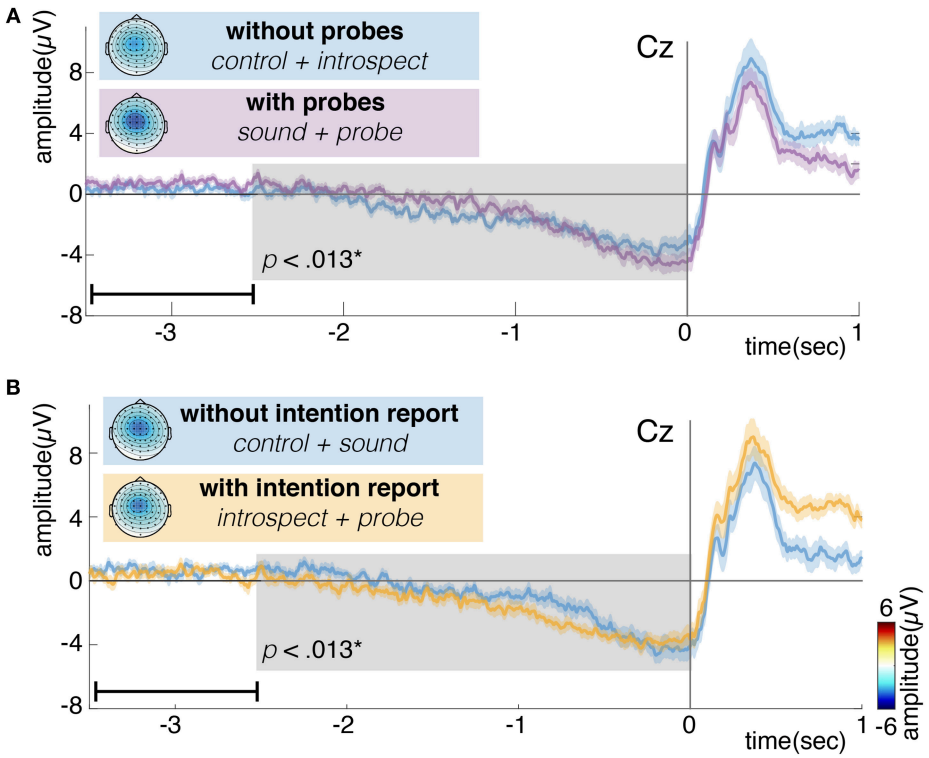


FIGURE 10 | (A) Grand average readiness potential at electrode Cz for conditions without (control + introspect) and with probes (sound + probe). **(B)** Grand average readiness potential at electrode Cz for conditions without (control + sound) and with intention reports (introspect + probe). Action onset is at time 0 and is indicated by a vertical line. The colored shade indicates the standard error across participants. The topoplots show the grand average EEG activity at each electrode, averaged across the last 0.5s prior to action onset. Significant differences ($p < 0.013$) are indicated by a gray box. Note, the data is baselined between 3.5 and 2.5 s prior to action, as indicated by a small horizontal line in the bottom left corner of each plot.

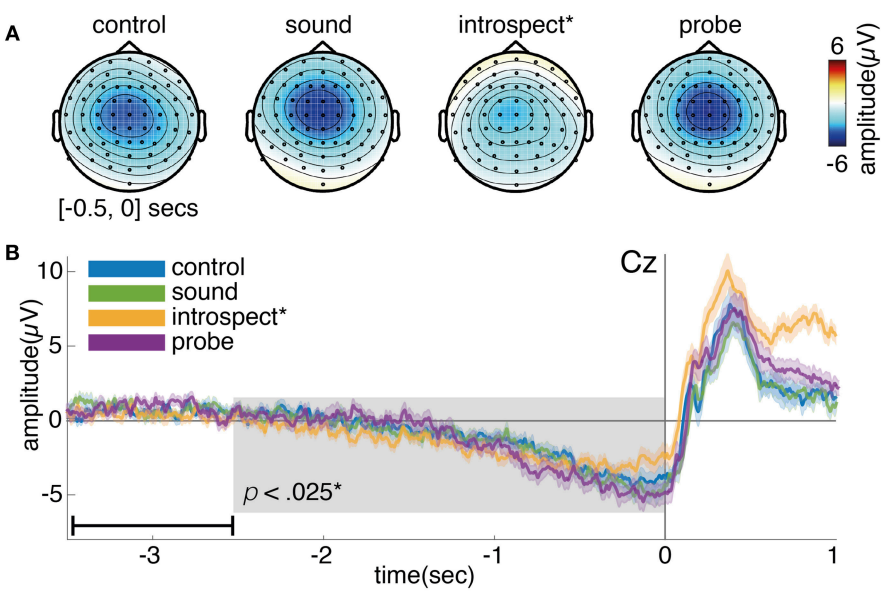


FIGURE 11 | (A) Topoplot showing the grand average EEG activity at each electrode, averaged across the last 0.5s prior to action onset. **(B)** Grand average of the readiness potential at electrode Cz. Each color shows the grand average RP of a single condition surrounded by two lines, indicating the standard error across participants. Action onset is at time 0 and is indicated by a vertical line. Note, the data is baselined between 3.5 and 2.5 s prior to action, as indicated by a small horizontal line in the bottom left corner of plot B. The introspect condition was found to differ significantly from all others ($p < 0.025$).

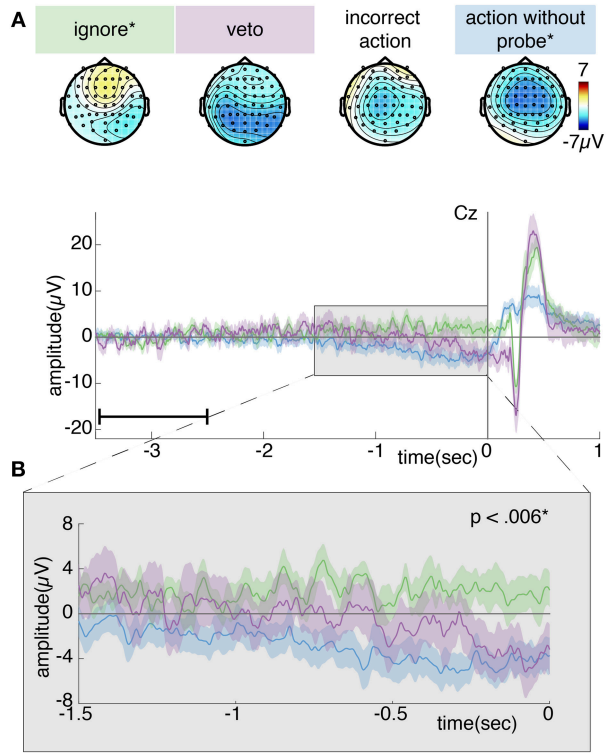


FIGURE 12 | (A) Topoplot showing the grand average EEG activity of the probe condition averaged across the last 0.5s prior to action (action-without-probe trials) or probe onset (ignore, veto, or incorrect action trials). **(B)** Grand average ERP at electrode Cz. Each color shows the grand average RP of a single trial type surrounded by the standard error across participants. For simplicity, we leave out the incorrect trials, since they closely resemble the action-without-probe trials. Time 0 corresponds to action (action-without-probe trials) or probe (ignore or veto) onset and is indicated by a vertical line. A significant difference between action-without-probe and ignore trials was found across –1.5 to 0 s. Note, the data is baselined between 3.5 and 2.5 s prior to action or probe onset, as indicated by a small horizontal line in the bottom left corner of plot **(B)**.

$df = 7, t = -5.4562, p = 0.001$) phase of the RP. No significant differences were found between the action-without-probe and incorrect action trials ($df = 7, t = -0.3285, p = 0.752$), action-without-probe and veto trials ($df = 7, t = -1.928, p = 0.095$) or veto and ignore trials ($df = 7, t = -1.014, p = 0.172$) for the early phase of the RP. Similarly, no significant differences were found between the action-without-probe and incorrect action trials ($df = 7, t = -0.4593, p = 0.670$), action-without-probe and veto trials ($df = 7, t = -1.969, p = 0.090$) or veto and ignore trials ($df = 7, t = -2.587, p = 0.018$) for the late phase of the RP.

Figure 12 also shows differences between action-without-probe, ignore, veto and incorrect action trials after time 0s. Note however, that time 0 s refers to action onset for action-without-probe trials only and to probe onset for ignore, veto and incorrect action trials. Therefore, any differences after time 0s likely reflect the presence of an action or probe (or the combination of the two for incorrect action trials). We do not investigate brain activity after time 0 s because any differences in brain processing related to action preparation will be contaminated by the brain response to probe presentation or action performance.

4.2.2. Alpha/Beta Event-Related Desynchronization

Significant main effects of the requirement of an introspection report and the presence of probes are found on the last 2.5 s of the RP ($N = 16$, $p < 0.013$, see **Figure 13**). **Figure 14** shows the grand average of the ERD across the alpha (8–12 Hz) and beta (13–30 Hz) bands at channel C3 for each condition. An ERD is visible in the average time-frequency spectrum prior and during action (time 0). The ERD seems to have its earliest onset at around 2 s prior to action onset. Judging from the topoplot provided in **Figure 14A**, the ERD seems enhanced in the introspect and especially probe conditions compared to the control and sound conditions. After action, an event-related synchonization is visible in the control, sound and probe conditions, whereas the ERD seems to continue after movement in the introspect condition. This continued ERD in the introspect condition is caused by the second button press that is required to answer the multiple-choice question on the subjective experience of the intention to act. Overall, the shape and timing of the observed ERD and ERS signatures confirm those of previous research (Pfurtscheller and Aranibar, 1979; Doyle et al., 2005). Individual *post-hoc* tests reveal significant differences in the ERD between the probe and sound, probe and control, and introspect and control conditions ($N = 16, p < 0.025$).

5. DISCUSSION

The probe method of Matsuhashi and Hallett (2008) promises to be a valuable addition to the clock method of Libet et al. (1983) in studying the timing of an intention to act. Here, it was put to test to verify its accuracy and quantify any potential concerns that might limit its usage. In a 2 × 2 within-subject design (1) the requirement of an introspection report and (2) the presence of an auditory probe were manipulated. In total, the experiment consisted of 4 conditions: *control*, *sound*, *introspect* and *probe*. In all conditions, participants were instructed to make a spontaneous self-paced right hand movement between 5 and 10 s after trial start. The *control* condition consisted solely of the performance of these voluntary actions. In addition, the *sound* and *probe* conditions contained (identical) auditory probes. Furthermore, the *introspect* and *probe* conditions required a report on the experienced intention to act. In the *introspect* condition, this report was provided by answering a multiple-choice question on the vividness of an experienced intention. In the *probe* condition, it was provided through the response to a probe: ignoring the probe or vetoing the intended action. In addition, 14 participants completed a questionnaire to get more insight into their subjective experience during the experiment. The effect of the two manipulated variables on the neural preparatory activity for action (RP and alpha/beta ERD), the awareness of an intention, and the performed actions was investigated.

The presence of probes had a significant effect on the observed mean action times. This effect was due to significant differences

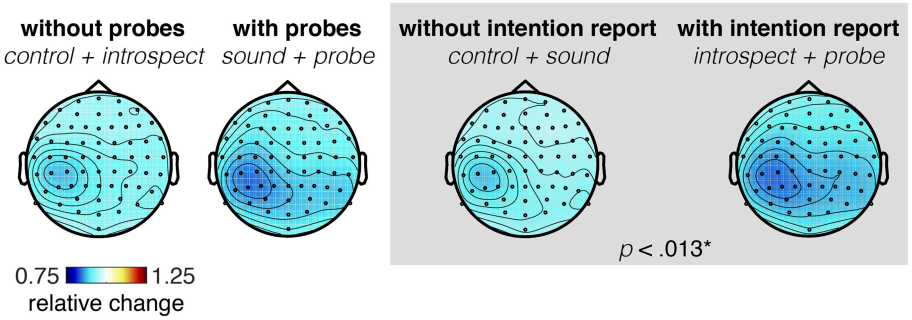


FIGURE 13 | Topoplots showing the grand average power across 8–30 Hz at each electrode, averaged across the last 0.5 s prior to action onset. Averages across conditions with (sound + probe) and without probes (control + introspect), and with (introspect + probe) and without intention reports (control + sound) are shown. A significant difference ($p < 0.013$) is found between conditions with and without intention reports.

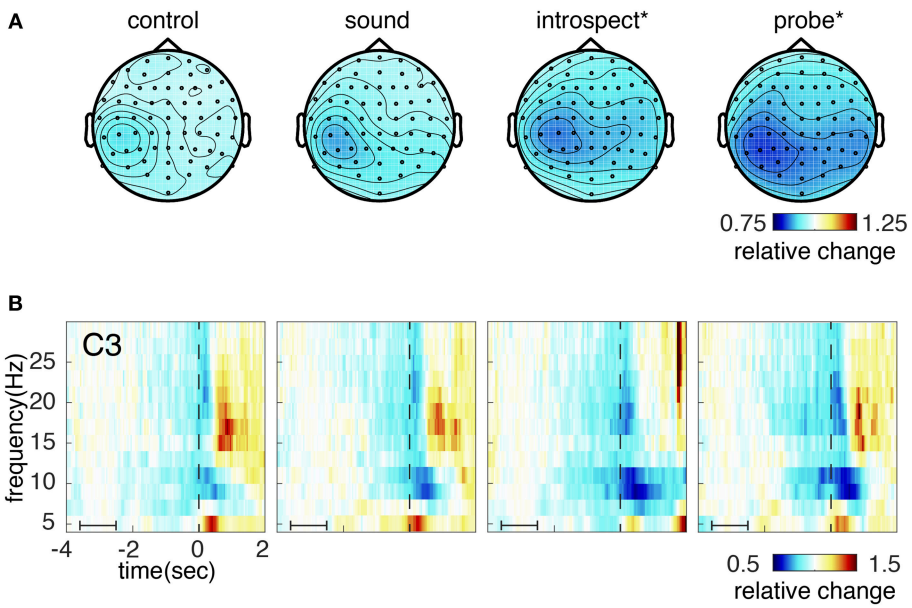


FIGURE 14 | (A) Topoplot showing the grand average power across 8-30Hz at each electrode, averaged across the last 0.5s prior to action onset. **(B)** Grand average of the event-related desynchronization across the alpha (8–1 2Hz) and beta (13–30 Hz) bands measured at electrode C3. Each spectogram corresponds to the grand average data of an individual condition: *control* (left), *sound* (left of middle), *introspect* (right of middle) and *probe* (right). The dotted vertical line indicates action onset (time 0). The data is baselined between 3.5 and 2.5 s prior to action, as indicated by the horizontal line in the bottom left corner.

in mean action time between the *sound* and *introspect*, *control* and *probe*, and *probe* and *introspect* conditions. Regardless of the requirement of an introspection report, the presence of probes seem to speed up the mean action time by roughly 0.3 s. One possible explanation for this observation is that the action times get tuned toward the probe distribution. The probe distribution is designed to be a shifted version of the action distribution. On average, this shift causes the probes to be presented slightly prior to the actions. This precedence of probes could bias the actions by speeding them up. Although a significant difference is found between the *control* and *probe* conditions, it was not found between the *control* and *sound* condition. On the one hand, this may be due to the fact that the sampling of actions in the *probe* condition slightly enhances the effect of the probes: slow actions are more likely to be turned into a veto response due to a probe. Because these slow actions will not be performed, they do not add to the mean action time. This enhances the difference between mean action onset in the *probe* compared to the *sound* condition. On the other hand, this result may also confirm hypothesis 5: veto influences action. Participants may act faster in the *probe* condition because they want to either explicitly (consciously) or implicitly (subconsciously) decrease the chance of having to perform a veto.

When specifically investigating the effect of the probes on the action times, the mean action times relative to probe onset of the *sound* condition were found to differ from those of the *introspect*

condition. However, the action times relative to probe onset did not differ significantly between the *sound* and *control* nor the *introspect* and *control* conditions, suggesting that the difference between the *sound* and *introspect* conditions are a combined effect from the relatively slower action times in the *introspect* condition and faster action times of the *sound* condition.

Although probes seem to speed up actions, they do not seem to speed up brain signals. In contrast, it is not the presence of probes, but the requirement of an introspection report that seems to influence brain signals: reducing the amplitude of the RP and increasing the desynchronization in the alpha and beta bands prior to action onset. Specifically, the RPs in the *introspect* condition have the lowest amplitude of all. This may suggest that the introspection requirement influences the threshold for action performance, making one more susceptible to small fluctuations in neural activity. Moreover, performing introspection may enhance the level of attention, which reduces the general level of alpha activity and could cause a relatively enhanced desynchronization in this range. These types of effects of introspection, i.e., tuning attention toward ones internal state, have been reported to induce changes in brain activity in the past (Lau et al., 2007).

Nine out of 12 participants who completed the questionnaire indicated that they found it difficult to assess whether or not they were intending to act upon probe presentation in the *probe* condition. Yet, vetos were performed in a specific time range across all participants: running from 1.4 until 0.65 s prior to action onset. Finding a significant descrease in the amount of ignored probes during this time range confirms the main behavioral findings of Matsuhashi and Hallett (2008). Moreover, within the probe condition, we found a significant difference in brain activity prior to an ignored probe and an action: we see a clear RP prior to action onset, but nothing like an RP prior to an ignored probe. This suggests that there is no neural preparatory activity for action present prior to an ignored probe, providing further evidence indicating that participants are indeed not experiencing an intention to act when they ignore a probe.

The average onset of the awareness of an intention to act found here and in previous research (Matsuhashi and Hallett, 2008; Verbaarschot et al., 2016) is much earlier than that found by studies using the clock method (Libet et al., 1983; Haggard and Eimer, 1999; Haggard et al., 2002; Sirigu et al., 2004; Banks and Isham, 2009; Bode et al., 2011; Fried et al., 2011; Soon et al., 2013; Jo et al., 2014; Douglas et al., 2015; Tabu et al., 2015; Alexander et al., 2016). This suggests that probes spread out the awareness of an intention to act. During a specific time range prior to action (starting around 1.4 s before), probes seem to facilitate the awareness of an intention. Within this time range, probes are consistently followed by a veto response across all participants. As argued previously (Uithol et al., 2014; Verbaarschot et al., 2016), these findings suggests that an intention to act reflects a dynamic process rather than a discrete mental state. The probe method is able to measure the awareness of intending during the earlier stages of this process, during which one seems susceptible to external stimuli. In contrast, the clock method seems to measure only its final stage, during which intentions become available for self-initiated report.

The probe distribution is designed to present the majority of probes in the last 3s prior to action onset, where awareness of an intention is most likely to occur (Matsuhashi and Hallett, 2008; Verbaarschot et al., 2016). The choice of probe distribution may co-determine the results: awareness of an intention to act starting earlier than 3s prior to action onset may be missed using our probe distribution. We believe our choice of probe distribution should be sufficient since our results as well as those of Matsuhashi and Hallett and Verbaarschot et al. show that probes are consistently followed by a veto response starting from 1.5s prior to action onset, and not earlier. If needed, the probe distribution could be extended in future research to investigate earlier awareness of an intention to act. If one would want to measure the timing of an intention to act across a larger time period, one would need to extend the experiment in order to retain a similar time resolution: a fixed amount of probes with small variations in timing provides a more detailed approximation of the timing of an intention to act during a limited time period, than the same amount of probes with big variations in timing across a larger time period.

Eleven out of 12 participants who completed the questionnaire indicated that they acted spontaneously throughout the experiment. Although participants were instructed to act whenever they experienced an intention to do so, only 45% of these intentions were reported as vivid and conscious across all participants in the *introspect* condition. The remaining intentions were either perceived as a vague feeling of wanting (34%) or without any conscious thought at all (21%). These results might be related to the type of task that participants were asked to perform: an arbitrary button press that is performed without any reason or consequence. In daily life, these actions are usually not preceded by a vivid intention. This type of task has been criticized in the past (Mele, 2010; Nachev and Hacker, 2014) and highlights the importance of designing an ecologically valid experiment that aims to measure meaningful actions (Mecacci and Haselager, 2015). In contrast to the clock method, the probe method can be used to reach this goal as it does not require constant introspection and can be used in combination with other stimuli. Additional stimuli can be used to create an ecologically valid experimental context in which actions can be made for a reason and have some consequence.

We used trial-by-trial color feedback rather than a sporadic verbal correction [as most probably used by Matsuhashi and Hallett (2008)] to prevent potential instructional differences between participants. In addition, the feedback was designed to provide an intuitive feeling of the time window of opportunity for acting without the need to keep track of time. However, the feedback may have enhanced the artificial nature of the experimental task and decreased the element of spontaneity in comparison to that of Matsuhashi and Hallett (2008). In a previous study we conducted a Matsuhashi style experiment without the use of color feedback (Verbaarschot et al., 2016). Participants had complete freedom to decide when to act and what action (a left or right hand button press) to perform. In our current study, we observed a similar action pattern to that of our previous Matsuhashi style experiment in terms of action mean and variance across all conditions. As we believe

that action variance increases with spontaneity, this suggests that the actions measured in this experiment are at least as spontaneous as those measured in our previous Matsuhashi type experiment that did not use any color feedback. Moreover, since participants acted around 5-7 s after trial start in the previous Matsuhashi experiment, the 5–10 s window of opportunity used in the current experiment does not seem to restrict the average participant's spontaneous actions in an obvious way (Verleger et al., 2016). In future research, we encourage the use of a more ecologically valid task that evokes the required action timing and pattern in a natural and intuitive way without the use of explicit feedback.

In summary, our test of the probe method has confirmed certain effects of the presence of probes and the requirement of an introspection report: probes speed up actions and introspection changes the neural preparation for action. However, we do not believe that these effects make the probe method an unsuitable alternative to the clock method to study the timing of intentions to act. The RP and alpha/beta ERD were clearly detectable in all experimental conditions, confirming previous findings on the neural preparation for a voluntary movement (Pfurtscheller and Aranibar, 1979; Libet et al., 1983; Shibasaki and Hallett, 2006). Moreover, vetos were performed consistently across participants from 1.4 to 0.65 s prior to action onset. Together with previous research, this time range has been confirmed in three independent investigations (Matsuhashi and Hallett, 2008; Verbaarschot et al., 2016). Probes do seem to affect the timing of actions, speeding them up a bit (about 300 ms on average compared to the *control* condition). But they do not seem to speed up or delay brain signals. The requirement of an introspection report does influence the brain signals. However, this effect is common to both the probe and clock methods, as they both require introspection reports. Moreover, the continuous introspection, as required by the clock method and mimicked by the *introspect* condition, seems to exacerbate these effects as shown by the lower amplitude RP and enhanced desynchronization in the alpha and beta bands prior to action onset.

We believe that the probe method provides a valuable addition to the clock method. The probe method can be used in combination with other tactile, visual, or even auditory (if the probe is easily detectable) stimuli, creating the possibility to embed it in a more realistic and ecologically valid experimental task. By including it in our repertoire, intentional actions can be studied in various experimental contexts. In contrast to the clock method, the probe method measures the awareness of an intention to act in real-time during action preparation. As such, it requires only sporadic introspection. Moreover, the probe method seems capable of measuring earlier stages of intending compared to the clock method. Depending on one's research objective, one might favor the clock or probe method over the other. When the amount of experimental time needs to be limited and one is interested in the onset of a reportable intention to act, one might best opt for the clock method. On the other hand, when devising a complex and ecologically valid experimental task and one is interested in the time period during which one is aware of an intention to act, the probe method seems the best way to go. With this overview and our current findings, we hope to encourage the use of Matsuhashi and Hallet's probe method in future research and extend the repertoire for experimentally studying intended action.

AUTHOR CONTRIBUTIONS

CV was involved in all stages of the research. In addition, she performed the data collection and analysis. All authors contributed equally to the experimental design. PH was mainly involved in the formation of research questions, the conceptual analysis and interpretation of the data. JF contributed mainly to the methods, data analysis and interpretation of subsequent results.

REFERENCES

Alexander, P., Schlegel, A., Sinnott-Armstrong, W., Roskies, A. L., Wheatley, T., and Tse, P. U. (2016). Readiness potentials driven by non-motoric processes. *Conscious. Cogn.* 39, 38–47.doi: 10.1016/j.concog.2015.11.011

Bai, O., Rathi, V., Lin, P., Huang, D., Battapady, H., Fei, D. Y., et al. (2011). Prediction of human voluntary movement before it occurs. *Clin. Neurophysiol.* 122, 364–372. doi: 10.1016/j.clinph.2010.07.010

Banks, W. P., and Isham, E. A. (2009). We infer rather than perceive the moment we decided to act. *Psychol. Sci.* 20, 17–21. doi: 10.1111/j.1467-9280.2008.02254.x

Bode, S., He, A. H., Soon, C. S., Trampel, R., Turner, R., and Haynes, J. D. (2011). Tracking the unconscious generation of free decisions using uitra-high field fmri. *PLoS ONE* 6:e21612. doi: 10.1371/journal.pone.0021612

Dominik, T., Dostál, D., Zielina, M., Šmahaj, J., Sedláčková, Z., and Procházka, R. (2018). Libet?s experiment: a complex replication. *Conscious. Cogn.* 65, 1–26. doi: 10.1016/j.concog.2018.07.004

Douglas, Z. H., Maniscalco, B., Hallett, M., Wassermann, E. M., and He, B. J. (2015). Modulating conscious movement intention by noninvasive brain stimulation and the underlying neural mechanisms. *J. Neurosci.* 35, 7239–7255. doi: 10.1523/JNEUROSCI.4894-14.2015

Doyle, L. M., Yarrow, K., and Brown, P. (2005). Lateralization of event-related beta desynchronization in the eeg during pre-cued reaction time tasks. *Clin. Neurophysiol.* 116, 1879–1888. doi: 10.1016/j.clinph.2005.03.017

Fried, I., Mukamel, R., and Kreiman, G. (2011). Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron* 69, 548–562. doi: 10.1016/j.neuron.2010.11.045

Gratton, G. (1998). Dealing with artifacts: The eog contamination of the event-related brain potential. *Behav. Res. Methods Instr. Comput.* 30, 44–53. doi: 10.3758/BF03209415

Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nat. Rev. Neurosci.* 9, 934–946. doi: 10.1038/nrn2497

Haggard, P. (2019). The neurocognitive bases of human volition. *Ann. Rev. Psychol.* 70, 9–28. doi: 10.1146/annurev-psych-010418-103348

Haggard, P., Clark, S., and Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nat. Neurosci.* 5:382. doi: 10.1038/nn827

Haggard, P., and Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Exp. Brain Res.* 126, 128–133. doi: 10.1007/s002210050722

Jo, H. G., Wittmann, M., Borghardt, T. L., Hinterberger, T., and Schmidt, S. (2014). First-person approaches in neuroscience of consciousness: brain

dynamics correlate with the intention to act. *Conscious. Cogn.* 26, 105–116. doi: 10.1016/j.concog.2014.03.004

Keller, I., and Heckhausen, H. (1990). Readiness potentials preceding spontaneous motor acts: voluntary vs. involuntary control. *Electroencephalogr. Clin. Neurophysiol.* 76, 351–361. doi: 10.1016/0013-4694(90)90036-J

Khalighinejad, N., Schurger, A., Desantis, A., Zmigrod, L., and Haggard, P. (2018). Precursor processes of human self-initiated action. *Neuroimage* 165, 35–47. doi: 10.1016/j.neuroimage.2017.09.057

Kornhuber, H. H., and Deecke, L. (1965). Hirnpotentialänderungen bei willkürbewegungen und passiven bewegungen des menschen: Bereitschaftspotential und reafferente potentiale. *Pflüg. Archiv. Euro. J. Physiol.* 284, 1–17. doi: 10.1007/BF00412364

Lau, H. C., Rogers, R. D., and Passingham, R. E. (2007). Manipulating the experienced onset of intention after action execution. *J. Cogn. Neurosci.* 19, 81–90. doi: 10.1162/jocn.2007.19.1.81

Lew, E., Chavarriaga, R., Silvoni, S., and Millán, Jdel R. (2012). Detection of self-paced reaching movement intention from eeg signals. *Front. Neuroeng.* 5:13. doi: 10.3389/fneng.2012.00013

Libet, B., Gleason, C. A., Wright, E. W., and Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential) the unconscious initiation of a freely voluntary act. *Brain* 106, 623–642. doi: 10.1093/brain/106.3.623

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of eeg-and meg-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

Matsuhashi, M., and Hallett, M. (2008). The timing of the conscious intention to move. *Euro. J. Neurosci.* 28, 2344–2351. doi: 10.1111/j.1460-9568.2008.06525.x

Mecacci, G., and Haselager, P. (2015). A reason to be free. *Neuroethics* 8, 327–334. doi: 10.1007/s12152-015-9241-8

Mele, A. R. (2010). Testing free will. *Neuroethics* 3, 161–172. doi: 10.1007/s12152-008-9027-3

Miller, J., Shepherdson, P., and Trevena, J. (2011). Effects of clock monitoring on electroencephalographic activity: is unconscious movement initiation an artifact of the clock? *Psychol. Sci.* 22, 103–109. doi: 10.1177/0956797610391100

Nachev, P., and Hacker, P. (2014). The neural antecedents to voluntary action: a conceptual analysis. *Cogn. Neurosci.* 5, 193–208. doi: 10.1080/17588928.2014.934215

Navon, D. (2014). How plausible is it that conscious control is illusory? *Am. J. Psychol.* 127, 147–155. doi: 10.5406/amerjpsyc.127.2.0147

Perrin, F., Pernier, J., Bertrand, O., and Echallier, J. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalogr. Clin. Neurophysiol.* 72, 184–187. doi: 10.1016/0013-4694(89)90180-6

Pfurtscheller, G., and Aranibar, A. (1979). Evaluation of event-related desynchronization (erd) preceding and following voluntary self-paced movement. *Electroencephalogr. Clin. Neurophysiol.* 46, 138–146. doi: 10.1016/0013-4694(79)90063-4

Rigoni, D., Kühn, S., Sartori, G., and Brass, M. (2011). Inducing disbelief in free will alters brain correlates of preconscious motor preparation: The brain minds whether we believe in free will or not. *Psychol. Sci.* 22, 613–618. doi: 10.1177/0956797611405680

Saigle, V., Dubljević, V., and Racine, E. (2018). The impact of a landmark neuroscience study on free will: a qualitative analysis of articles using libet and colleagues' methods. *AJOB Neurosci.* 9, 29–41. doi: 10.1080/21507740.2018.1425756

Schlegel, A., Alexander, P., Sinnott-Armstrong, W., Roskies, A., Peter, U. T., and Wheatley, T. (2013). Barking up the wrong free: readiness potentials reflect processes independent of conscious will. *Exper. Brain Res.* 229, 329–335. doi: 10.1007/s00221-013-3479-3

Schneider, L., Houdayer, E., Bai, O., and Hallett, M. (2013). What we think before a voluntary movement. *J. Cogn. Neurosci.* 25, 822–829. doi: 10.1162/jocn_a_00360

Schultze-Kraft, M., Birman, D., Rusconi, M., Allefeld, C., Görgen, K., Dähne, S., et al. (2016). The point of no return in vetoing self-initiated movements. *Proc. Natl. Acad. Sci. U.S.A.* 113, 1080–1085. doi: 10.1073/pnas.1513569112

Schurger, A., Sitt, J. D., and Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2904–E2913. doi: 10.1073/pnas.1210467109

Shibasaki, H., and Hallett, M. (2006). What is the bereitschaftspotential? *Clin. Neurophysiol.* 117, 2341–2356. doi: 10.1016/j.clinph.2006.04.025

Sirigu, A., Daprati, E., Ciancia, S., Giraux, P., Nighoghossian, N., Posada, A., et al. (2004). Altered awareness of voluntary action after damage to the parietal cortex. *Nat. Neurosci.* 7:80. doi: 10.1038/nn1160

Soon, C. S., Brass, M., Heinze, H.-J., and Haynes, J. D. (2008). Unconscious determinants of free decisions in the human brain. *Nat. Neurosci.* 11, 543–545. doi: 10.1038/nn.2112

Soon, C. S., He, A. H., Bode, S., and Haynes, J. D. (2013). Predicting free choices for abstract intentions. *Proc. Natl. Acad. Sci.* 110, 6217–6222. doi: 10.1073/pnas.1212218110

Tabu, H., Aso, T., Matsuhashi, M., Ueki, Y., Takahashi, R., Fukuyama, H., et al. (2015). Parkinson's disease patients showed delayed awareness of motor intention. *Neurosci. Res.* 95, 74–77. doi: 10.1016/j.neures.2015.01.012

Uithol, S., Burnston, D. C., and Haselager, P. (2014). Why we may not find intentions in the brain. *Neuropsychologia* 56, 129–139. doi: 10.1016/j.neuropsychologia.2014.01.010

Verbaarschot, C., Haselager, P., and Farquhar, J. (2016). Detecting traces of consciousness in the process of intending to act. *Exper. Brain Res.* 234, 1945–1956. doi: 10.1007/s00221-016-4600-1

Verleger, R., Haake, M., Baur, A., Śmigasiewicz, K., et al. (2016). Time to move again: does the bereitschaftspotential covary with demands on internal timing? *Front. Hum. Neurosci.* 10:642. doi: 10.3389/fnhum.2016.00642

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometr. Bullet.* 1, 80–83. doi: 10.2307/3001968

Wisniewski, D., Goschke, T., and Haynes, J. D. (2016). Similar coding of freely chosen and externally cued intentions in a fronto-parietal network. *Neuroimage* 134, 450–458. doi: 10.1016/j.neuroimage.2016.04.044

Wohlschläger, A., Haggard, P., Gesierich, B., and Prinz, W. (2003). The perceived onset time of self-and other-generated actions. *Psychol. Sci.* 14, 586–591. doi: 10.1046/j.0956-7976.2003.psci_1469.x

Wolpe, N., and Rowe, J. B. (2014). Beyond the "urge to move": objective measures for the study of agency in the post-libet era. *Front. Hum. Neurosci.* 8:450. doi: 10.3389/fnhum.2014.00450

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

frontiers in Human Neuroscience

REVIEW
published: 19 March 2019
doi: 10.3389/fnhum.2019.00057

A Review of Psychophysiological Measures to Assess Cognitive States in Real-World Driving

Monika Lohani[1]*, *Brennan R. Payne*[2] *and David L. Strayer*[2]

[1] *Department of Educational Psychology, University of Utah, Salt Lake City, UT, United States,* [2] *Department of Psychology, University of Utah, Salt Lake City, UT, United States*

As driving functions become increasingly automated, motorists run the risk of becoming cognitively removed from the driving process. Psychophysiological measures may provide added value not captured through behavioral or self-report measures alone. This paper provides a selective review of the psychophysiological measures that can be utilized to assess cognitive states in real-world driving environments. First, the importance of psychophysiological measures within the context of traffic safety is discussed. Next, the most commonly used physiology-based indices of cognitive states are considered as potential candidates relevant for driving research. These include: electroencephalography and event-related potentials, optical imaging, heart rate and heart rate variability, blood pressure, skin conductance, electromyography, thermal imaging, and pupillometry. For each of these measures, an overview is provided, followed by a discussion of the methods for measuring it in a driving context. Drawing from recent empirical driving and psychophysiology research, the relative strengths and limitations of each measure are discussed to highlight each measures' unique value. Challenges and recommendations for valid and reliable quantification from lab to (less predictable) real-world driving settings are considered. Finally, we discuss measures that may be better candidates for a near real-time assessment of motorists' cognitive states that can be utilized in applied settings outside the lab. This review synthesizes the literature on in-vehicle psychophysiological measures to advance the development of effective human-machine driving interfaces and driver support systems.

Keywords: psychophysiology, cognition, driving, traffic safety, real-world

OPEN ACCESS

Edited by:
Bruce Mehler, Massachusetts Institute of Technology, United States

Reviewed by:
Joost De Winter, Delft University of Technology, Netherlands
Mickael Causse, National Higher School of Aeronautics and Space, France
Dick De Waard, University of Groningen, Netherlands
Edmund Wascher, Leibniz Research Centre for Working Environment and Human Factors (IfADo), Germany

***Correspondence:**
Monika Lohani
monika.lohani@utah.edu

Received: *01 May 2018*
Accepted: *01 February 2019*
Published: *19 March 2019*

Citation:
Lohani M, Payne BR and Strayer DL (2019) A Review of Psychophysiological Measures to Assess Cognitive States in Real-World Driving. Front. Hum. Neurosci. 13:57. doi: 10.3389/fnhum.2019.00057

THE IMPORTANCE OF PSYCHOPHYSIOLOGICAL MEASURES IN TRAFFIC SAFETY

Suboptimal level of cognitive functioning (e.g., inattention, drowsiness) is a key cause of traffic accidents and poor driving performance. According to Traffic Safety Culture Index, 87.5% of drivers identify distracted driving to be a greater concern today than in past years and 87.9% perceive drowsiness as a threat to their safety (AAA Foundation for Traffic Safety, 2018). Traffic safety researchers are constantly working on methods to improve driving performance by assessing cognitive states, such as drivers' workload, inattention, and fatigue. One way to improve the assessment of covert cognitive states is to adopt a multi-method approach to measure changes in central and peripheral nervous system functioning in order to sense near-real time information about cognitive states of motorists. Such assessments of internal states can also promote the development

of Advanced Driver Assistance Systems (ADAS) that can predict and augment risky driving behavior.

Why Adopt Psychophysiological Measures?

Cognitive states can be assessed using subjective, behavioral, and physiological measures (Mauss and Robinson, 2009; Strayer et al., 2015; Lohani et al., 2018). Subjective measures can be limiting if the assessment is disruptive to the real-time task (i.e., primary task intrusion, see O'Donnell and Eggemeier, 1986). More importantly, humans may not always be accurate in making judgements about their cognitive states (Schmidt et al., 2009). Motorists can be inaccurate in making judgments about their internal and cognitive states (such as their attention, workload, and drowsiness levels). For instance, motorists were inaccurate at self-assessments of vigilance (Schmidt et al., 2009); even though objective physiological indicators (e.g., heart rate, EEG, and ERPs) suggested poor vigilance levels at the end of a 3-h drive, participants self-reported improved vigilance instead (Schmidt et al., 2009). Such misjudgments in assessment of cognitive states suggest that objective measures are required to assess and augment human behavior in order to reduce risk for traffic safety. While behavioral measures (such as head movement detection to assess distraction) are also useful, given the intent of this review, we will focus on physiological measures. Accuracy in detecting cognitive workload has been found to significantly increase when physiological data was utilized (Lenneman and Backs, 2009, 2010; Solovey et al., 2014; Borghini et al., 2015; Yang et al., 2016). Some work has also found that physiological measures were sensitive to variations in cognitive load during secondary tasks while behavioral driving measures like steering wheel reversals and velocity (Belyusar et al., 2015) and lane-keeping measures (Lenneman and Backs, 2009) were not. Unlike behavioral measures (e.g., verbal and facial behavior), many physiological measures are not under voluntary control of motorists. Moreover, cognitive states such as mental workload are a multi-faceted and dynamic concept and self-report alone cannot be used to operationalize it, but multiple measures (e.g., performance and physiology) are warranted (de Waard and Lewis-Evans, 2014). Thus, inclusion of physiological data can complement and extend behavioral metrics and improve assessments of motorists' state-level changes in cognition (Brookhuis and de Waard, 1993; Mehler et al., 2012).

As automation is likely to become more prevalent over time, real-time monitoring behaviors required by motorists may decline as they are less involved in the driving process. This is a critical reason why non-behavior-based metrics will become more relevant to incorporate into our understanding of the motorists' cognitive states. Moreover, distracted motorists of a self-driving vehicle compared to manually driving motorists take longer to gain control of the driving task once automation deactivates (Vogelpohl et al., 2018). Intelligent driving assistance systems should be capable of reliably sensing and assessing distraction and drowsiness levels of motorists to be able to augment safe-driving conditions. Building reliable systems to be able to predict decreased levels of vigilance or dangerous levels of fatigue, drowsiness, or workload could help augment them in a timely manner (Balters et al., 2018).

Cognition in Dynamic Real-World Driving Contexts

In general, psychophysiological measures can be used to assess degree of arousal or activation (Mauss and Robinson, 2009). Importantly, multiple psychological constructs can influence variations in psychophysiological measures. For instance, heart rate, skin conductance, and electrical activity of the brain are sensitive to many psychological constructs experienced by motorists, such as workload, drowsiness, stress, etc. In the past years, important contributions have reviewed the literature on specific cognitive states, such as workload (Borghini et al., 2014; Costa et al., 2017), distraction (Matthews et al., 2019), drowsiness (Sahayadhas et al., 2012; Borghini et al., 2014), and stress (Rastgoo et al., 2018) in driving research. These reviews provide an understanding of physiological outcomes that can explain variations in specific constructs based on carefully manipulated and well-controlled designs. Unlike highly controlled lab-based settings, where a single construct (e.g., workload) can be successfully manipulated and its effect on psychophysiological measures examined, real-world settings are more dynamic and complex.

In a real-world setting, the net resulting cognitive state of a motorist is a combination of variation among several interrelated constructs (e.g., attention allocation, stress, workload, fatigue). Broadly speaking, the net cognitive state of a motorist, composed of variation among these many dimensions, can be classified along an arousal-spectrum ranging from lower-arousal and passive states, to a state of optimal performance, to a hyper-aroused or over-active state. Indeed, this concept is not new; Yerkes and Dodson (1908) established strong non-linear relationships between arousal-level and performance, and such relationships have since been well-established across many human performance domains (Hebb, 1955; Broadhurst, 1959; Wekselblatt and Niell, 2015). Although these ideas are not new, there has been a recent resurgence in a formal understanding of arousal-performance relationships, including an expanded understanding of the underlying neuromodulatory systems involved in regulating task engagement and optimal performance (e.g., the adaptive-gain control theory, Aston-Jones and Cohen, 2005). Given the recent increase in understanding of the mapping between physiological indices of arousal and human performance in the lab, such models serve as a clear starting point in delineating the predictive capacity of psychophysiological measures for understanding cognitive states and human performance in the vehicle.

For instance, low-arousal states relevant to the driving task can be driven by a combination of psychological constructs including low workload, reduced stress, and high drowsiness. On the other hand, an over-aroused state could be due to a combination of high workload and high stress in the presence of low drowsiness. Similarly, other combinations of constructs can also lead to changes in general arousal states as well. Given the likely dynamic interplay among these interrelated constructs in applied settings,

the current review focuses on psychophysiological measures that can be utilized to capture motorists' states in real-world driving settings. Indeed, one major applied goal of this work is to be able to accurately capture the dynamic and highly variable changes in arousal that occur in ecologically valid driving settings, a goal that is critical for building accurate predictive models (Yarkoni and Westfall, 2017) of individual motorist's states and future driving performance.

Specifically, there are two novel contributions of this review. First, instead of focusing on a selective construct and related measures of interest, the goal of this current review is to focus on psychophysiological measures that may have the potential to be adopted in real-world and applied settings to measure state level variations in motorists. The paper provides a broad but selective review of a number of psychophysiological measures that we believe show the greatest promise in their utilization to assess low-arousal vs. over-arousal (passive vs. over-active) states in real-world driving environments. The most commonly used physiology-based measures of cognitive states are considered as potential candidates relevant for driving research. The following physiological measures are reviewed (see section "Psychophysiological Measures to Assess Cognitive States" and **Tables 1**, **2**) in assessing arousal state in real-world driving research: electroencephalography and event-related potentials, optical imaging, heart rate, and heart rate variability, blood pressure, skin conductance, electromyography, thermal imaging, and pupillometry. As reviewed in classical contributions by Cacioppo et al. (Cacioppo and Tassinary, 1990; Cacioppo et al., 2007), inference of unique psychological constructs based on physiological indices (one-to-one relation) is still unresolved and is not the aim of this review (see further discussion in section "Research Applicability in Real-World Settings"). However, we discuss how multiple measures (that are sensitive to several interrelated internal states) may be combined to delineate net resulting changes across multiple inter-related cognitive state-level variations. Second, for each measure, we make the distinction between useful research measures and practical measures for real-world application (see section Research Applicability in Real-World Settings and **Table 2**). Throughout, we have tried to highlight the practical relevance of measures in the driving context. Although this review focuses primarily on on-road and simulated driving contexts, when relevant, we have also drawn research from related contexts (traffic operators, pilots, or ship navigators) to more thoroughly characterize each measure.

PSYCHOPHYSIOLOGICAL MEASURES TO ASSESS COGNITIVE STATES

Electroencephalogram (EEG) and Event-Related Potentials (ERP)

EEG Quantification

The *EEG* is a record of both oscillatory and aperiodic brain electrical activity. Neural activity (largely post-synaptic potentials) from multiple simultaneous generators propagate throughout the brain and skull and summate at a distance, where voltages can be measured relatively non-invasively via electrodes placed on the scalp. The dominant sources of scalp-recorded EEG come from cortical pyramidal cells arranged in the columnar organization of the cortex (Nunez and Srinivasan, 2006). Pyramidal cells are the most numerous cortical excitatory cell type and play a critical role in advanced cognitive functions (Spruston, 2008). The laminar organization of the cortex results in cortical pyramidal cells following an *open-field* alignment with a consistent orientation that is perpendicular to the skull, such that their post-synaptic potentials can summate at a distance. Importantly, EEG allows for a high temporal resolution (millisecond) and direct record of neural activity. This detailed temporal resolution also allows for a decomposition of the time-domain EEG signal into spectral information via Fourier analysis, allowing for an examination of oscillatory activity in canonical frequency bands (e.g., alpha, ∼8–12 Hz; theta, ∼4–7 Hz), which have been related to specific neurocognitive functions. For instance, mental workload increases theta power and reduce alpha power activity (Mun et al., 2017), whereas fatigue increases alpha power (Käthner et al., 2014). Moreover, the development of novel computational techniques for analyzing spectral activity has promoted a wide range of new tools for probing ongoing neural dynamics during human cognition via EEG; such as cross-frequency coupling, phase coupling (Cohen, 2011), independent component analysis (Dasari et al., 2017), and neighborhood component analysis (Lim et al., 2018). In addition, more traditional analyses of transient neural activity that is tied to specific perceptual, motor, or cognitive events can be gleaned from continuous EEG, via the calculation of event-related brain potentials.

ERP Quantification

ERPs are electrophysiological responses that are consistently linked in time with specific sensory, cognitive, or motor events. They are derived from the continuously recorded EEG by time-aligning epochs of EEG relative to an event of interest, such as a stimulus onset or a participant's response and averaging many of these similar EEG segments to reveal activity that is time and phase locked to the event. Such discrete events can be added in the experimental design, e.g., every time a participant responded to a secondary task while driving. The logic of this approach is that systematic activity that is locked in time and space to some specific activity will remain in the averaged ERP waveform, whereas activity that is not time- and phase-locked will average to zero with a large enough number of trials (Luck and Kappenman, 2012). The resulting ERP waveform is plotted as voltage over time at a given set of electrodes. ERP topography can also be examined, showing the distribution of activity over the entire space within a particular time-window. A major benefit of ERPs is that the waveform has characteristic *components*, stereotyped features of the ERP with specific eliciting conditions. ERP components are defined empirically by a combination of their polarity, timing, scalp distribution, and sensitivity to task manipulations.

Extensive work has characterized and validated specific ERP components with respect to their associations with specific cognitive and neural processes (e.g., Fabiani et al., 2007; Luck and Kappenman, 2012; Mun et al., 2017). Cognitive

TABLE 1 | Overview of relationships between arousal state and physiological indices in real-world driving.

Measure	Under-arousal state	Over-arousal state
Electroencephalogram	• Increased alpha due to increase in drowsiness and attentional withdrawal[b] • Changes in theta and delta activity related to transition to fatigue	• Increase in theta activity due to mental workload • Alpha activity suppression due to workload
Event related potential	• Reduced ERP amplitudes with fatigue, time on task, and lower vigilance over time while driving	• Also, reduced ERP amplitude to driving relevant stimuli under high workload
Optical imaging for cerebral flow	• A decrease in cerebral oxygenation with fatigue and drowsiness	• An increased concentration of oxygenated hemoglobin and a decreased concentration of deoxygenation with mental workload and stress
Heart rate and Heart rate variability[b]	• Decrease in heart rate with drowsiness, decrease in vigilance, use of self-driving technology • Increase in HRV indices (e.g., RMSSD) with drowsiness, fatigue, and disengagement	• Increase in heart rate with mental workload and stress • Decrease in HRV with workload, stress, and vigilance
Blood Pressure[b]	• Decrease in blood pressure relative to baseline with fatigue and drowsiness	• Increase in systolic BP with workload and stress
Electrodermal activity	• Lower EDA relative to baseline activity range with drowsiness	• EDA increase with workload, stress, lower trust in automation, and anxiety
Electromyography	• Decreases in mean and median power frequency of EMG due to decline in muscle activities and fatigue [a]	• High muscle activity relative to baseline with stress
Thermal imaging	• Temperature around baseline levels [a]	• Higher task difficulty increases forehead temperature and decreases nose temperature and thus an increase in the difference between forehead and nose temperatures
Pupillometry	• Decreases in average pupil diameter with drowsiness • Increases in standard deviation of pupil diameter	• Increases in pupil diameter with cognitive load • Decreases in standard deviation of pupil diameter

[a] *Limited findings available.*
[b] *Mixed findings reported.*

demands can modulate several ERP components, such as P3 (discussed below; Käthner et al., 2014), mismatch negativity (MMN is a negative ERP component sensitive to pre-attentive information processing; Wanyan et al., 2018), and late positive potentials amplitude (a later ERP component like P6 that is related to attentional allocation similar to P3; Mun et al., 2017). The P3 component is associated with attentional and memory processes required to detect any changes in incoming stimuli-related information (Polich, 2007). The canonical P3 has two distinct but related components – the P3a and P3b (see Polich, 2007 for a review). The P3a, with an anterior distribution, is associated with novel stimulus-driven attentional processing or orienting responses. The P3b, with a centro-posterior distribution, is associated with task-relevant stimulus-driven attentional, decision making, and subsequent memory processing (Polich, 2007). Both components have been used in driving research. Recent work has also examined how neural indices (as measured by both P3 ERP components) are associated with subjective workload (as measured by NASA-TLX) and how this covariation is influenced by cognitive effort (Yakobi, 2018). Novel techniques (such as intra-block averaging of ERP amplitudes; Horat et al., 2016) can enable robust electrophysiological measurement of cognitive demands over time. Thus, ERPs are an attractive measure for studying cognitive states and performance in driving contexts.

EEG/ERP in Driving Context

EEG and ERPs have a long history in the study of the neural indices of cognitive effort and attention allocation in both laboratory and applied settings. EEG is perhaps one of the most widely used neurophysiological methods to study driving behavior. Several frequencies (e.g., power in alpha frequency band) and time (e.g., P3) domain indices can reliably measure changes in cognitive demands (Käthner et al., 2014). This makes EEG is viable measure for applied driving settings.

Over-arousal in driving context

Over-aroused states, such as increased workload while driving can be indexed by decreases in alpha power and increases in theta power (Borghini et al., 2014; Käthner et al., 2014). A recent study found alpha band power to be higher during the relaxed condition compared with the engaged condition in an autonomous driving setting (Zander et al., 2017). This highlights the sensitivity of alpha power to internal factors such as attentional engagement. In addition to internal factors, external factors (such as task load and time on task) can also influence alpha and theta power bands in opposite directions (Wascher et al., 2018). For instance, a decrease in task load and time on task led to an increase in relative alpha power, but a decrease in theta power (Getzmann et al., 2018; Wascher et al., 2018). To account for both power bands, past work has also used a ratio of frontal theta and parietal alpha power spectral density to operationalize workload in pilots (Borghini et al., 2015). This ratio approach may be relevant for driving research as well, however this has been a point of debate, as discussed shortly.

The application of known ERP indicators of attentional workload (and their eliciting tasks) can be successfully translated into the driving domain as well. One of the most commonly adopted components in driving research is the P3b (Brookhuis and de Waard, 2010; Solís-Marcos and Kircher, 2018). Mental

TABLE 2 | Tentative framework for considering the research applicability of these measures in lab and real-world settings.

Measure	Lab-setting	Real-world	Advantages	Disadvantages
Electroencephalogram	High	High to medium	• High temporal resolution • Direct measure of neural activity	• Contact sensors • Longer setup time
Event related potential	High	Medium[a]	• Same benefits as EEG • Well-characterized components (e.g., P3) • High temporal resolution	• Same disadvantages as EEG • Generally needs higher number of trails and post-processing • Needs time-locking event • Interpretation of amplitude is very context specific
Optical Imaging for Cerebral Flow	High	Low to medium[a]	• Higher spatial resolution • Feasible in naturalistic settings with technical advancements	• Lower temporal resolution (e.g., fNIRS) • Need systematic investigation/replication
Heart Rate/Heart Rate Variability	High	High	• Higher signal-to-noise ratio (SNR) • Easy to collect	• Very sensitive to artifacts • Sensitive to variation in respiration
Blood Pressure	High	Medium[a]	• Reliable • Higher SNR	• Limitations of equipment; can disrupt driving task
Electrodermal activity	High	High	• Sympathetic activity • Easy to collect	• Lagged response • Not all individuals show EDA response
Electromyography	High	Low[a]	• Reliable • High temporal resolution	• Slightly longer setup time • Sensitive to movement • Lower SNR
Thermal Imaging	High	Medium[a]	• Low setup time • Promising technology • Non-contact	• Need systematic investigation/replication
Pupillometry	High	Low[a]	• Non-contact • Quick setup time	• Signal strongly sensitive to variable lighting conditions (pupillary light reflex)

[a] *Limited findings available.*

workload can be indexed by increases in P3b latencies (Ying et al., 2011) and amplitude (Strayer and Drews, 2007). For example, Strayer and Drews (2007) examined the amplitude of the P3b time-locked to the onset of a pace break light under single-task driving conditions or dual-tasking via cell-phone–induced distraction. Drawing on basic experimental work that has shown that the P3b is sensitive to the degree of attention allocated to a task (e.g., Sirevaag et al., 1989), they also showed that cell-phone induced distraction resulted in reduced P3b amplitudes to brake lights. Similar effects have been observed in comparing the workload of "single-task" driving in laboratory simulator vs. real-life driving contexts, where for example, the diversion of attention to other concurrent activities in the vehicle result in additional attentional demands in real-world driving (Strayer et al., 2015).

A recent study compared mental workload due to increased information processing demands consumed by in-vehicle information systems (Solís-Marcos and Kircher, 2018). They found both P3b and N1 latencies and amplitudes to be sensitive to cognitive demands of processing additional in-vehicle information systems. For instance, P3b amplitudes decreased with additional information processing related tasks (Solís-Marcos and Kircher, 2018). P3a amplitude was also found to decrease with additional task-related load (Getzmann et al., 2018). High mental workload has been associated with increased latencies in MMN during driving (Ying et al., 2011) and also increased frontal MMN in flight simulation tasks (Wanyan et al., 2018), however a recent study did not find workload to influence MMN amplitudes (Getzmann et al., 2018). Future work will help clarify sensitivity of MMN in driving research.

Under-arousal in driving context

Extensive work has focused on electrophysiological indicators of under-arousal via EEG. A substantial number of papers have implicated changes in alpha amplitude during fatigued driving (e.g., Schier, 2000; Jensen and Mazaheri, 2010; Simon et al., 2011; Zhao et al., 2012; Borghini et al., 2014; Jagannath and Balasubramanian, 2014; Arnau et al., 2017; Brouwer et al., 2017), such that fatigued driving is associated with increased alpha activity. However, other work has challenged these alpha power links with fatigue and claim that alpha power changes may be due to the decreases in task-demands and visual input during monotonous driving tasks and not due to decline in cognitive processing abilities (Wascher et al., 2014). Increases in relative alpha band power with increased time on task, easier driving route, and lower control of driving situations, which suggested that relative alpha power increases imply attentional withdrawal and not fatigue (Wascher et al., 2014, 2018). Wascher et al. (2014, 2018) have argued that mid-frontal theta activity may be a more appropriate neural marker of cognitive-control related processes in driving than occipital alpha activity. Low task load is associated with relatively reduced theta activity, which suggests that theta activity is sensitive to declines in cognitive processing ability. Instead of alpha activity, Wascher et al. (2014, 2018) recommend that indices of oscillatory synchronization (e.g., inter-trial phase clustering) and ERPs (such as P3a) are more reliable and valid

indices of changes in cognitive state associated with mental fatigue. For instance, time on task (Wascher et al., 2014), fatigue (Massar et al., 2010), and decreases in vigilance over time (Schmidt et al., 2009) were found to reduce P3a amplitude while driving. Similarly, mind-wandering during driving is associated with a reduction in P3a amplitude (Baldwin et al., 2017). One other study found both P3a and P3b components' amplitudes were reduced due to driving-related fatigue (Guoping and Zhang, 2009). These findings show that ERP components could be utilized to detect variations in neurophysiological arousal due to interrelated cognitive constructs in driving contexts.

Some researchers have argued that LF/HF ratios (e.g., frontal theta/beta) are potential biomarkers for attentional control, and have established some evidence that such measures have good psychometric properties, for e.g., test-retest reliability (Putman et al., 2014; Angelidis et al., 2016). Decreases in beta power (e.g., Zhao et al., 2012; Jagannath and Balasubramanian, 2014) have been found, along with changes in theta and delta activity as markers related to transition to fatigue. This has led some researchers to propose spectral ratio indices (e.g., alpha/beta; Eoh et al., 2005; Wang et al., 2018), as biomarkers of alertness. However, ratio indices have also been criticized for being an inadequate method because it combines frequency bands with distinct topographic specificity that change differently over time (Wascher et al., 2014). There is existing criticism of this ratio approach, especially in driving research (Wascher et al., 2018), and more broadly, researchers in cognitive electrophysiology have been moving away from such highly constrained "band-based" approaches given their lack of replicability across studies. Alternatively, researchers have increasingly endorsed methods that allow for broad-band assessment of spectral dynamics (e.g., 1/f scaling, Voytek and Knight, 2015) and methods that can address narrow-band dynamics without a priori selection of frequency (e.g., cluster-based permutation testing in time-frequency data; Maris and Oostenveld, 2007). Other recent work has used EEG-based detection algorithms to detect fatigue and drowsiness (Li et al., 2017; Morales et al., 2017; Belakhdar et al., 2018; Gao et al., 2018; Wei et al., 2018). However, other work reported no additional benefit of utilizing EEG measures in drowsiness and fatigue detection in sleep deprivation contexts (Perrier et al., 2016; Liang et al., 2017). Another line of work has aimed to apply machine-learning techniques to brain computing interfaces in order to classify states of drowsiness and fatigue in real-time (e.g., Lin et al., 2005; Correa et al., 2014). Recent work has also shown data filtering and processing techniques such as artifact subspace reconstruction and independent component analysis could be utilized for "online" processing of EEG data collected while driving in order to attenuate movement-and noise-related artifacts (Krol et al., 2017). Together, these findings suggest that EEG and ERPs can be utilized as objective techniques to assess state-level variations in cognitive demands.

Practical Considerations

There are a number of important considerations when applying EEG indices to real-world driving environments. Typical EEG artifacts arising from muscle-and-eye movements (de Waard, 1996; Zander et al., 2017), impedance shifts, environmental line (60 Hz) noise, and other complications are potentially amplified in real-world environments. As such, real-time monitoring of good quality EEG signals is critical for effective data collection. The commercial introduction of high-impedance systems with active electrodes and small electrically shielded mobile EEG amplifiers has spawned a large increase in real-world EEG applications. Many of these systems are capable of high density (<128 channel) recording, but it is critical for the researcher to decide whether and to what degree an increase in the number of channels may result in a decrease in the quality of the recorded EEG (Luck and Kappenman, 2012). Importantly, the well-understood limitations of the spatial resolution of EEG limit the utility of high-density recording in ecologically valid environments (e.g., where measurement of EEG sensors co-localized in 3D space on a single-subject basis may be unfeasible). Moreover, with increasing channel density comes increases in the likelihood for poorly recorded or poorly monitored channels during recording. As such, if source-localization of underlying EEG/ERP generators is not a primary aim of the methodology (and we expect, in most applied cases it would not be), researchers may wish to record from a smaller density (e.g., 32 channels or fewer), at the benefit of better monitoring of data quality throughout the experiment.

On the theoretical side—researchers in human factors automotive research should carefully consider the linking hypotheses between specific electrophysiological indicators (e.g., P3b ERP amplitude, alpha power increases) and their purported cognitive interpretations. The ERP literature has a massive basic literature in which specific components have been very well-characterized relative to their eliciting conditions and underlying cognitive interpretations (Luck and Kappenman, 2012). One such example was reviewed earlier on characterizing the P3b under different states of distraction during driving. Limited work (e.g., Strayer et al., 2015) has attempted to examine ERP components in naturalistic settings. In future work, inventive approaches can be validated to use task-related responses or behaviors (such as eye-blink potentials or frequent vs. infrequent vehicle cues) as discrete events that can be recorded to estimate ERP components in real-world settings. At the same time, such characterizations in the spectral domain are not as clearly developed to date. However, this is changing, as basic research in cognitive electrophysiology shifts toward a more complete understanding of oscillatory mechanisms underlying human perception and cognition (e.g., Kahana, 2006), involving development in standardized analysis methods (Cohen, 2011), careful experimental characterization of specific oscillatory markers (e.g., alpha phase and perception, Mathewson et al., 2009; midline frontal theta and conflict resolution; Cavanagh and Frank, 2014), and the development of neurophysiologically guided models (Jensen and Mazaheri, 2010; Voytek and Knight, 2015). We expect that such development of basic research findings in cognitive electrophysiology will be a great asset in future applied research in contexts such as driving.

Optical Imaging for Cerebral Blood Flow

Optical Imaging Quantification

Optical imaging methods allow for the visualization of the interaction of photons with tissues (Villringer et al., 1993). In recent years, there has been a rapid advancement in the application of non-invasive optical imaging methods such as functional near infrared spectroscopy (fNIRS) to study human brain and cognitive functioning. fNIRS is a neuroimaging method based on the principles of near-infrared spectroscopy, which was originally developed in humans for investigating clinical features of brain functioning (e.g., cerebral oxygenation; Jobsis, 1977). These principles have been extended to measure local changes in cerebral hemodynamic activity that can be used to infer information on the underlying neural activity due to neurovascular coupling, following similar logic to the Blood Oxygen Level Dependent (BOLD) signal in functional magnetic resonance imaging. NIR (700–1,000 nm) light is able to penetrate several centimeters through the skull and into brain tissue, allowing for non-invasive measurement of certain optical properties of cortical tissue. For example, changes in the concentration of oxy- and deoxy-hemoglobin can be measured via NIRS because oxy- and deoxy-hemoglobin have distinct absorption spectra that correspond to the different coloration of arterial and venous blood (Grinvald et al., 1986). These absorption characteristics make it possible to use a spectroscopic approach to measure changes in the concentration of oxy- and deoxy- hemoglobin as a function of neural activity, for example during cognitive task performance. In typical optical imaging systems, optical fibers, called optodes or sources, carry NIR light to the scalp while other optical fibers, called detectors, collect the photons as they emerge from the scalp. Each source–detector pair is a single channel. Multi-channel and wearable fNIRS systems have become commercially available with diverse montages capable of measuring brain activity across the entire scalp.

Optical Imaging for Cerebral Blood Flow in Driving Context

The application of fNIRS in driving research is in its infancy. Nevertheless, a number of interesting demonstrations of the utility of fNIRS for studying over-arousal states such as driver workload have emerged (e.g., Tsunashima and Yanagisawa, 2009; Liu et al., 2012, 2016; Sibi et al., 2016). For example, increases in oxygenated hemoglobin have been reported during simulated driving tasks under cognitive load compared to control conditions (Liu et al., 2012). A recent study (Unni et al., 2017) utilized fNIRS in a naturalistic driving simulator while doing a secondary task (modified version of 0–4 back). They found systematic increases in bilateral inferior frontal and temporo-occipital brain regions with increments in workload. Another study reported that fNIRS could be used to differentiate between low vs. high workload (n-back task) related hemodynamic activity in the prefrontal cortex while motorists drove in a realistic driving simulator (Herff et al., 2017). Furthermore, fNIRS have been used to monitor pilot's task engagement and working memory load in real-time (Gateau et al., 2015). On a related note, fNIRS have been found sensitive to increase in task difficulty in flight simulators (Causse et al., 2017) as indicated by an increased concentration of oxygenated hemoglobin and a decreased deoxygenated hemoglobin.

Other work has investigated effects of under-arousal related states with fNIRS. Research has related decreases in hemodynamic measures of cerebral oxygenation with fatigue in simulated driving (Li et al., 2009), and findings have been extended into actual highway driving (Yoshino et al., 2013). An increase in fatigue can be indexed by a decrease in cerebral oxygenation and mental stress can be indexed by an increase in cerebral oxygenation. Tsunashima and Yanagisawa (2009) examined changes in prefrontal activity via multi-channel frontal fNIRS systems in driving with and without adaptive cruise control. Their findings revealed substantial decreases in prefrontal activity when participants drove with adaptive cruise control relative to without, which was correlated with perceived workload (via the NASA-TLX). Similar decreases in activation of prefrontal cortex (lower cognitive load associated with drowsiness) were reported while participants monitored a simulated autonomous car driving task relative to higher prefrontal cortex activation during manual driving task (Sibi et al., 2016). Such findings indicate that optical imaging for cerebral blood flow is a valuable tool for assessing performance and neural efficiency in well-controlled realistic driving contexts.

Practical Considerations

One important limitation of fNIRS is that, because it relies on the measurement of absorption properties of light as a function of vascular changes in the brain, its temporal resolution is limited by the time-course of hemodynamic activity (on the order of seconds). In contrast, the development of recent 'fast' optical imaging methods, such as the event-related optical signal (EROS; Gratton and Fabiani, 2001, 2003), which measures scattering properties of light as a function of changes in neural activity, have a much higher temporal resolution (on the order of milliseconds). Although applications of this method in human factors research is sparse, fast optical imaging methods have growing promise. While the spatial resolution of optical imaging methods is higher than EEG, such spatial inference is constrained by the penetration depth of NIR light, which reaches only a few cm from the scalp surface. Therefore, imaging of activity from deep cortical and subcortical sources (beyond the outer cortical mantle) is limited. Recent work has also employed wearable fNIRS systems (Piper et al., 2014; McKendrick et al., 2016; Le et al., 2018) and simultaneous collection of fNIRS and EEG (Kassab et al., 2018), which can enable real-world monitoring in ecologically valid settings.

Heart Rate (HR) and Heart Rate Variability (HRV)

Heart Activity Quantification

Heart rate (in beats per minute or bpm) is the number of heartbeats in 1 min (Jennings et al., 1981). Electrocardiography (ECG) is a well-established method to record the electrical activity of the heart. In psychophysiology, a lead II configuration (i.e., placing the negative electrode in the region of right collar bone, the ground near the left collar bone, and the positive lead over the lower left ribcage, or functionally similar variant) is

commonly used to be able to record electrical activity of the heart via research grade equipment. A single heart beat wave in an ECG signal shows changes in electrical potentials (referred to as the P, Q, R, S, & T components and together they are referred to as the *QRS complex*, for review please see Berntson et al., 2007). The R component (one for each heart beat) is due to ventricular depolarization and for a lead II configuration, it has a larger magnitude and a sharper inflection than the rest of the components making it easily detectable. While heart rate is a count of beat per minute, *heart period* (also called inter-beat-interval) is the time in milliseconds between successive R spikes (Berntson et al., 2007). Heart rate is generally derived by converting mean heart period (in milliseconds) to heart rate (in beats per minute), see Berntson et al. (2007).

Heart data can also be collected via other technique including photoelectric plethysmography (PPG) and photoplethysmography imaging (PPGI). PPG technique includes use of a photocell (such as an infrared light-emitting diode) placed over an area of tissue with blood capillaries that is easily accessible (e.g., finger or ear lobe). Energy emitted from an infrared source passes through the tissue and reflects off the tissue. Changes in blood volume (due to heart beats) in an area can thus be assessed by the amount of light that was reflected back to the photodetector, and thus forms the basis of estimating heart beats (Berntson et al., 2007; Laborde et al., 2017). A similar concept is used in "wearables" which have photo-emitters and detectors placed on a convenient location (e.g., wrists and earlobes) making them easy to wear and collect data from them (Byrom et al., 2018; van Gent et al., 2018). This idea is used in vehicles with photo-emitters and detectors placed on the steering wheels, which allow collecting heart data (heart rate, HRV, and blood volume pulse) while driving. Another advancement in PPG is a contactless measurement technique called PPGI that detects color changes (e.g., the forehead area) in a video due to blood perfusions (Blöcher et al., 2017). Instead of photodiodes used in PPG, PPGI uses detector arrays in cameras to collect image sequences that contain information about bio-signals (e.g., blood volume pulse and respiration). Image and signal processing methods are utilized for beat-to-beat heart rate estimation (Blöcher et al., 2017; Madan et al., 2018).

On a related note, established guidelines for heart beat detection processing, with recommended parameters to derive heart rate and heart rate variability are provided in Jennings et al. (1981), Berntson et al. (2007), and Shaffer and Ginsberg (2017). Custom and open-source software has also been developed to automatically detect R peaks to calculate heart beats. As is true for most physiological measures, data should be visually checked to inspect the ECG data for artifacts and irregularities. Artifacts can be introduced in these data due to numerous reasons (such as motorists' excessive motion, sneezing and coughing, and irregular heartbeats) any of which can disrupt the ECG measurement or directly impact normal heart-beat patterns. Visual inspection helps insure that the heart beats are correctly marked by the detection software and physiologically improbable values are detected and then corrected.

HRV is variability in the time intervals of adjacent heartbeats (Berntson et al., 2007; Shaffer and Ginsberg, 2017). HRV can be derived from ECG data over a period of time ranging from short intervals (∼1–5 min) up to longer intervals (∼24 h). HRV metrics can be roughly categorized as falling under time-domain, frequency-domain, or non-linear measures of HRV (for a review see Shaffer and Ginsberg, 2017). Time domain-based parameters calculate the variations in heart beat intervals, such as standard deviation of R-R intervals (SDRR), percentage of successive R-R intervals that differ by more than 50 ms (pNN50), and root mean square of successive R-R intervals (RMSSD). A few time-domain parameters also represent geometric shape of R-R interval distributions, such as the HRV triangular index (i.e., plotting the integral of the ratio of RR interval density histogram by its height) and the baseline width of the RR intervals histogram (TINN), for details see Shaffer and Ginsberg (2017). Frequency-domain based measures transform the beat-to-beat variations in heart beat (R-R intervals) into frequency power bands via Fourier analysis (Task Force of the European Society of Cardiology, 1996). The most commonly used frequency-domain methods are low- and high-frequency power. A low-frequency (LF) power is the energy of heart rate oscillations in a lower-frequency (0.04–0.15 Hz) band. Similarly, high-frequency (HF) power is the energy of heart rate oscillations in a higher-frequency (0.15–0.4 Hz) band (Task Force of the European Society of Cardiology, 1996; Shaffer and Ginsberg, 2017). A peak in these frequency bands can also be calculated, which is an estimate of the peak frequency in the specific frequency band. Non-linear measures of HRV are useful in capturing the unpredictability and dynamic nature of heart rate time-series data (Shaffer and Ginsberg, 2017). Common measures include fitting an elliptical-shape to represent non-linear HRV and calculating approximate entropy (ApEn) and sample entropy (SmpEn), which characterize the complex pattern of time-series heart data (Shaffer and Ginsberg, 2017). Detailed discussions can be found elsewhere (Task Force of the European Society of Cardiology, 1996; Berntson et al., 2007; Laborde et al., 2017; Shaffer and Ginsberg, 2017).

HR/HRV in Driving Context

Over-arousal in driving context

Heart rate is a commonly measured index of physiological arousal in response to changes in driving demands. One of the most studied over-aroused cognitive states is workload. Numerous studies have examined changes in heart rate as a function of workload (Lenneman and Backs, 2009, 2010; Mehler et al., 2012; Heine et al., 2017). Heart rate was also found to increase while performing visual and auditory dual-tasks relative to single-task of driving in a simulator (Lenneman and Backs, 2009). Similarly, heart rate has been shown to be incrementally higher for systematically more difficult auditory dual-tasks while driving in a simulator (Mehler et al., 2009) as well as while driving on-road (Reimer et al., 2009). These findings of an incremental change in heart have been replicated in younger-aged (20–29 years old), middle-aged (40–49 years old), and older-aged (60–69 years old) adults (Mehler et al., 2012). Thus, heart rate increases with workload due to cognitive demand (Lenneman and Backs, 2009; Mehler et al., 2012; Ruscio et al., 2017; Hidalgo-Muñoz et al., 2018; c.f., Engström et al., 2005). Other efforts have also been made to utilize rhythmic and morphological parameters

of a heart activity to explore mental workload. A recent study examined the influence of mental workload (due to a secondary task) on morphological parameters from ECG while completing a lane change task (Heine et al., 2017). They found that a combination of derived HR and HRV features (such as mean HR, RMSSD, pNN50, etc.) could be extracted from ECG data that could distinguish between workload levels and suggest that a combination of ECG features can be used to detect mental workload (for details see Heine et al., 2017).

Relative to HR, a fewer number of studies have examined HRV, especially in a systematic manner. HRV decreases with increasing task demands (Luque-Casado et al., 2016). HRV has been found to be sensitive to variations in attention levels while driving that may not be necessarily evident in driving performance (Lenneman and Backs, 2009) and thus HRV can have more sensitivity than behavioral measures. LF- and HF-HRV power bands are influenced by driving task (Zhao et al., 2012; Tozman et al., 2015; Wang et al., 2018). A study (Tozman et al., 2015) compared effect of demand levels (boredom, average demand, and high demand) on HRV in a driving simulator. Both LF- and HF-HRV varied for all the three conditions. High task demands reduced both LF-HRV and HF-HRV (Tozman et al., 2015). Some work has indicated that stress-inducing real-world driving tasks lead to increased heart rate and decreased SDNN, RMSSD, pNN50 (Lee et al., 2007). HRV also varies with workload experienced by drivers during simulated driving (Zhao et al., 2012; Heine et al., 2017; Hidalgo-Muñoz et al., 2018) and on-road driving (Lee et al., 2007). In addition, HRV variations due to cognitive workload have also been found in city traffic operators (Fallahi et al., 2016) and unmanned aerial vehicles operators (Jasper et al., 2016). HRV is sensitive to workload increases due to vigilance and situational awareness demands of the task (Saus et al., 2001; Stuiver et al., 2014; Jasper et al., 2016). However, at least one study (Shakouri et al., 2018) found no variation in heart rate variability metrics (RMSSD, LF, HF, and LF/HF ratio) as a function of higher traffic density while driving in a simulator, even though variations in subjective workload were found.

Under-arousal in driving context

HR and HRV are also sensitive to low-arousal states, such as vigilance and drowsiness. Decreases in vigilance over the course of a 3-h continuous driving task were indexed by a significant drop in heart rate over time (Schmidt et al., 2009). Drowsiness experienced in car drivers and aircraft pilots can also be associated with decreases in HR (Borghini et al., 2014). A recent on-road study (Biondi et al., 2018) found that driving a Tesla in semi-automated mode (e.g., autopilot) led to a lower heart rate relative to manual driving on a freeway. Another study found heart rate was sensitive to activity of the Adaptive Cruise Control (ACC) technology (Brouwer et al., 2017). Heart rate increased when ACC decelerated more suddenly compared to instances when the car decelerated more gradually (Brouwer et al., 2017). These findings suggest that heart rate is a sensitive measure that can assess cognitive processing pertaining to advanced technology in semi-autonomous vehicles.

Other studies have found that LF-HRV and HF-HRV vary with fatigue (Liang et al., 2009; Sugie et al., 2016). A recent study (Wang et al., 2018) found that changes in fatigue levels while driving can be represented by non-linear measures of HRV (e.g., sample entropy). Variations in drowsiness levels can also impact HRV (Noda et al., 2015; Piotrowski and Szypulska, 2017). Another recent study found that variations in HRV (TINN and RMSSD) was higher when participants drove a vehicle in automated mode relative to the manual mode (Biondi et al., 2018). Perhaps, drowsiness and a lack of engagement in the driving task during automated mode may have led to a higher HRV. HRV and blink rates have also been shown to assess sleep onset (Noda et al., 2015). HRV-based assessment algorithms can be used for early detection of fatigue and drowsiness to augment attention and performance (Patel et al., 2011; Zhao et al., 2012; Abe et al., 2016; Vicente et al., 2016).

Practical Considerations

Heart rate and its variability are inexpensive and reliable measures that are relatively easy to record with research-quality equipment that meets recommended guidelines (Task Force of the European Society of Cardiology, 1996). It has good signal to noise ratio as well (R-R peaks can be detected even in very noisy environments). Consequently, it is also not difficult to collect in lab as well as in unpredictable field studies, especially with the availability of mobile data recording systems. However, these advantages can also lead to misuse of this methodology. Great attention to the data collection and processing are required to have meaningful data. Skin preparation (e.g., cleaning with alcohol wipes) before electrode placement and signal monitoring to collect good quality data can drastically reduce post-processing (e.g., Berntson et al., 2007). Participants should be comfortably positioned to avoid physiologically induced changes in heart rate such as altered breathing rate due to postural adjustments. Body movements should be minimized and accounted for as such movements can add noise and also add movement-related heart rate changes. Effective data cleaning to remove artifacts and noise are a must, otherwise heart data will be uninterpretable.

Some recording devices do not utilize the traditional QRS complex from an ECG to calculate HR and HRV. For example, PPG uses a photoelectric sensor that estimates changes in blood volume to calculate HR. There are a few methodological challenges that should be considered before adopting such PPG-based systems. PPG records a lagged cardiac response further away from the heart (e.g., from fingers and earlobes). Unlike ECG based estimates that have a sharp spike for the R component, PPG-based methods instead show a less pronounced curved peak of the blood volume pulse signal, which makes accurate and automatic detection of heart period relatively more difficult (Laborde et al., 2017). Moreover, ECG-based estimates of HR and HRV are recommended for more reliable results because it allows visual inspection and artifact correction of heart data. Such methodological differences between PPG and ECG can explain why PPG and ECG findings are comparable during rest, but are not comparable during stress, for example (Schäfer and Vagedes, 2013).

On a related note, commercialized equipment meant for exercise and fitness tracking fail to meet established guidelines for heart data collection and processing (e.g., minimum

sampling rate and access to raw data for necessary artifact correction methods), which are necessary to make meaningful interpretations (see Berntson et al., 2007; Quintana et al., 2016; Shaffer and Ginsberg, 2017). Similarly, smartphone camera-based assessments have methodological challenges, including very poor sampling rate, illumination variation (due to confounds like weather and time of day), poor signal-to-noise ratio, and motion-related artifacts that can lead to inaccurate interpretations (Laborde et al., 2017; cf., Nowara et al., 2018; van Gent et al., 2018). Ensuring the validity and inter-device variability of wearables (which utilize a PPG-based or camera-based HR system) with an established ECG-based equipment is a necessary step to be able to validate data collected from wearables. However, most commercialized equipment has not been validated in such a manner (Quintana et al., 2016). Without this critical validation step, data collected from commercialized non-research grade equipment does not have convergent validity and should be discouraged by the scientific community until such standards are met. While innovation is critical to be able to collect psychophysiological data in real-world settings, careful adoption and cross-checks with existing gold standards are necessary to make meaningful progress in the adoption of these technologies in real-world driving research.

Moreover, HF-HRV has been found to be impacted by parasympathetic nervous system, however, LF-HRV is influenced by both sympathetic and parasympathetic nervous systems (Berntson et al., 2007; Laborde et al., 2017). Thus, LF-HRV should not be described as a metric of sympathetic activity, but instead be interpreted as a mixture of sympathetic and parasympathetic influences. On a related note, the LF/HF ratio has been a controversial metric as it assumes that LF is due to sympathetic activity while HF is due to parasympathetic (Billman, 2013). The LF/HF ratio was originally based on 24 h recordings, while shorter duration recordings (even 5 min long) have also been calculated. The duration of recording (e.g., 5 min vs. 24 h) can also lead to uncorrelated findings and some metrics are better for short term recordings than others (Shaffer and Ginsberg, 2017).

Another metric we would like to highlight is heart period. Heart rate and heart period have been used interchangeably, however in some instances heart period may be a better choice. Even though, heart rate is more commonly used metric, use of heart period instead of heart rate is recommended measure of autonomic activity because heart period changes more linearly over time (Quigley and Berntson, 1996; Berntson et al., 2007). Heart period should specially be used when comparing changes in heart activity due to experimental manipulation or due to between group differences for short time periods. Further information on heart activity related metrics can be found in detailed reviews (Jennings et al., 1981; Task Force of the European Society of Cardiology, 1996; Berntson et al., 2007; Laborde et al., 2017; Shaffer and Ginsberg, 2017).

Not all heart-based metrices may be sensitive to the variations in cognitive state during driving task. For instance, a study compared several commonly used metrices for HR and HRV cognitive workload during highway driving (Mehler et al., 2011). While HR was robust in differentiating between cognitive workload in single vs. dual tasks, HRV indices were less robust (e.g., smaller effect sizes). A few HRV indices varied with workload (RMSSD, SDSD, and LF power), however others (SDNN, NN50, pNN50, HF power, and LF/HF) did not significantly differ with workload (Mehler et al., 2011). These findings suggest that depending upon the task, certain indices may be more sensitive to variation in cognitive state than other indices that may be less robust.

In addition, researchers should consider other contextual factors that may vary across participants and may confound study interpretations. A confounding factor that can potentially bias HF-HRV comparisons between conditions of interest is differences in respiration (Grossman, 1992; Berntson et al., 2007; Laborde et al., 2017). Respiration related-parameters should be accounted for by using them as covariates with such HRV indices (for a detailed discussion, see Berntson et al., 2007; Laborde et al., 2017). Similarly, other factors may impact HR/HRV, including task characteristics and motorists' state (relaxation, engagement, and motivation) and activities (smoking and posture). For instance, HRV may increase over time if the task becomes less difficult over time, which may put motorists in a more relaxed state (Jasper et al., 2016). Similarly, HRV may also increase over time with disengagement or demotivation to perform a difficult task (Jasper et al., 2016). Careful consideration of contextual factors will afford accurate and reliable measurement of HR/HRV indices in applied driving settings.

Blood Pressure (BP)

BP Quantification

BP (in millimeters of mercury, also written as mmHg) is the force exerted against the walls of the blood vessels (Shapiro et al., 1996; Berntson et al., 2007). Depending upon the stage of the dynamic cardiac cycle, BP differs from lowest to highest levels. During a single cardiac cycle, *diastolic BP* is the lowest level of arterial pressure when the heart is filled with blood and *systolic BP* is relatively the highest level of arterial pressure (Shapiro et al., 1996; Berntson et al., 2007). As invasive methods to record BP require additional safeguards and equipment, most psychophysiology research studies focus on non-invasive approaches to record blood pressure. Three relatively non-invasive methods are auscultatory or oscillometric methods, arterial tonometry, or the volume-clamp methods (see for details, Berntson et al., 2007). The most common method is auscultatory measurement, which records the sounds of blood flow by placing a cuff on the upper arm and a stethoscope placed over the brachial artery to identify the systolic and diastolic blood pressure (Shapiro et al., 1996; Berntson et al., 2007). Physiological arousal during mentally effortful situations leads to greater vasoconstriction and cardiovascular reactivity evidenced by increased heart rate and blood pressure and decreased heart rate variability (Lundberg et al., 1994; Ottaviani et al., 2016). BP increases with psychological stress (Ottaviani et al., 2016) and is correlated with self-reported stress (Lundberg et al., 1994). However, cognitive workload may not reliably influence BP (ElKomy et al., 2017).

BP in Driving Context

Limited research has examined over- and under-arousal via BP in driving contexts. Systolic BP and BP variability have been found to increase while driving in simulated high traffic conditions that had high workload demands (Stuiver et al., 2014). Fatigue was also associated with a decrease in systolic BP and HR (Liang et al., 2009). However, other studies have not found a reliable effect of stress on BP (Simonson et al., 1968; Littler et al., 1973; Lee et al., 2007). One study found no significant change in BP from beginning to end of the drive with a short period of arterial pressure changes during events such as overtaking that returned to baseline (Littler et al., 1973). BP was also not found to vary in an on-road stressful driving task speed in a simulator even though HRV parameters were significantly impacted (Lee et al., 2007).

Nevertheless, BP is a very useful measure to understand the factors that impact driving performance. One clear example of this comes from a simulator-based study investigating aggressive driving behavior in irregular traffic flow and under time pressure (Drews et al., 2012). Irregular traffic patterns were not found to impact BP. However, male drivers who were under time pressure to drive faster in order to receive a monetary incentive, had elevated systolic BP compared to females under time pressure or compared to male drivers who were not under time pressure. In fact, females did not show any elevated blood pressure under time pressure (Drews et al., 2012). These findings suggest that individual difference factors such as sex differences and motivation to drive aggressively may impact driving behavior and associated physiological signals. Other studies have shown that trait-level variation in BP (such as a history of high BP i.e., hypertension) is an important measure to capture health and age-related impact on driving performance in vulnerable older populations (Lyman et al., 2001; Siren et al., 2004). A 5-year longitudinal study that examined the effect of urban bus driving on BP found that the number of hours driven per week predicted higher diastolic BP (Johansson et al., 2012), suggesting that there are cumulative effects of cognitive demands and stress of continuous driving.

Practical Considerations

While heart-rate was reported to rapidly change in response to car racing, BP was "less responsive" (Simonson et al., 1968). Other studies have found that BP does not change significantly during on-road driving (Littler et al., 1973; Lee et al., 2007). A few BP recording-related reasons could play a role. BP can rapidly change over time so multiple readings are recommended for a more accurate estimate. However, a limiting factor is the BP equipment. The pressure from a cuff worn by the responder can become uncomfortable and disruptive within a few minutes. Continuous reliable BP measurement (especially via volume-clamp) is uncomfortable, distracting, and potentially disruptive to driving. This limits the frequency of samples that could be collected, which are about 1 reading per minute. Also, the BP recordings are sensitive to movement so in an on-road study, it is less feasible to accurately record multiple BP reading from participants while drivers are actively involved in the driving process. While some alternative methods to record blood pressure (e.g., plethysmography) may be available, methodological issues similar to those discussed in recording heart activity apply to BP as well and it is crucial to evade poor quality unreliable equipment. In sum, BP provides valuable insights about vulnerable states of the drivers, however, in a real-world driving context, methodological concerns can limit reliable data collection. Much future work is required to be able to measure reliable and non-invasive BP activity.

Electrodermal Activity (EDA)

EDA Quantification

EDA, previously known as galvanic skin response, is a change in electrical potentials of the skin that can be used to make interpretations about the psychological phenomena of the responder (Boucsein et al., 2012). EDA can be measured via exosomatic or endosomatic techniques. Exosomatic techniques—a more commonly used method used in applied research—apply a small current through a pair of electrodes and then measure electrical resistance (or its reciprocal, i.e., conductance) from the skin. Because the current is kept constant, it is possible to measure changes in the voltage between the electrodes that will vary directly with changes in skin resistance, following Ohm's lab (see Dawson et al., 2007 for a technical review). Endosomatic techniques measure passive changes in intrinsic electrical activity without application of an external current. For details on EDA recording techniques, see Fowles (1986), Dawson et al. (2007), and Boucsein et al. (2012). Higher EDA is indicative of physiological arousal due to increased sympathetic autonomic nervous activity (Dawson et al., 2007; Lohani and Isaacowitz, 2014). EDA is sensitive to physiological reactivity and many other factors, such as respiration and mental effort (Dawson et al., 2007). Commonly derived EDA metrics (Dawson et al., 2007; Boucsein et al., 2012) include slowly varying tonic level of electrical conductivity (skin conductance level; SCL) and phasic increase in magnitude electrical conductance in response to an unexpected or relevant event (skin conductance response; SCR). Non-linear EDA metrics that can differentiate between increased cognitive load vs. recovery phases of stressors have been identified as well (Visnovcova et al., 2016).

EDA in Driving Context

In driving research, systematic variation in several arousal-related constructs can impact EDA. Most commonly investigated is cognitive workload. SCL is higher during increased workload in dual-task relative to single-task driving (Mehler et al., 2012). A systematic investigation of workload increments in one on-road driving study (Mehler et al., 2012) found a systematic increase in SCL as a function of three levels of auditory workload secondary tasks relative to single driving task for young, middle, and older age groups. These findings suggest that SCL can be used to index workload levels in driving context. High SCR has also been found to increase with workload experienced by motorists while driving on difficult road types that required avoiding more traffic and making more decisions (Schneegass et al., 2013). A recent study reported SCR amplitude increased with cognitive load due to dual-task driving (Ruscio et al., 2017). Additional workload experienced due to texting and navigation (Seo et al., 2017) and

speeding (Kajiwara, 2014) while simulated driving was also found to increase EDA.

EDA also varies with other physiological arousal-related constructs. EDA based indices can be used to detect stressful events during driving (Affanni et al., 2018). A recent study utilized feature extraction and discrimination processing techniques to classify EDA data into low, medium, vs. high stress levels with about 82% recognition rate (Liu and Du, 2018). Another recent study found higher SCLs when participants drove a simulated vehicle in autonomous mode compared to manual mode (Morris et al., 2017). Higher skin conductance levels could be indicative of lower levels of trust in the autonomous mode than manual mode. State anxiety during simulated driving was also found to be associated with SCL (Barnard and Chapman, 2018). Another recent study found that relative to sleepiness, higher skin conductance levels are found during wakefulness, effects which are indicative of comparatively higher sympathetic activity (Schmidt et al., 2017).

Practical Considerations

In driving contexts, EDA is shown to vary due to many cognitive states, such as workload, stress, anxiety, sleepiness, all of which are influenced by sympathetic nervous system activity. This allows the use of EDA in assessment of various psychological phenomena (Dawson et al., 2007). Therefore, caution should be exercised while interpreting changes in EDA in an applied and less-controlled setting as it is sensitive to not one, but many psychological variables. In the driving context, careful choice of filters to remove artifacts (Affanni et al., 2018) and identification of cognition-related features (Chen et al., 2017; Liu and Du, 2018) that have been successfully implemented could be utilized to improve accuracy and detection. One disadvantage of EDA is that it has a slower response (lag of 1–3 s) after the stimulus has occurred (Dawson et al., 2007). In instances when near-real time physiological responses need to be detected, EDA may be relatively slower (than cardiovascular measures). Another point to consider is that, similar to other physiological measures, not all individuals have the expected skin conductance response (Dawson et al., 2007). This is another reason to avoid reliance on a single measure, but multiple channels, to capture the psychological phenomena of interest.

Electromyography (EMG)

EMG Quantification

EMG is used to measure the electrical activity generated by muscle fibers (Fridlund and Cacioppo, 1986; van Boxtel, 2001). Surface EMG is captured by placing small surface electrodes on specific muscles of interest, which is then digitized and amplified to record muscle activity (Fridlund and Cacioppo, 1986). Numerous features can be extracted from the EMG signals. Root mean square of the signal (in microvolts) is a recommended and commonly reported EMG signal amplitude (Fridlund and Cacioppo, 1986). Other commonly assessed statistical features are peak spectral density, peak amplitude, and peak frequency. A specific muscle's activity can provide insights into the psychological processes underplay. For instance, the smile muscle (or zygomaticus major) and the frown muscle (or corrugator supercilii) have been used a lot in emotion research to identify positive and negative behavioral expressions. For example, more frown muscle activation can be an index of negative behavioral expressions (Lohani and Isaacowitz, 2014; Lohani et al., 2018). Psychological processes (e.g., stress) can lead to sympathetic nervous system activity (Lundberg et al., 1994), which can elicit muscular tension. Researchers have studied muscular activations under controlled conditions to index mental processes (Lundberg et al., 1994; Wijsman et al., 2013; Luijcks et al., 2014). Applied driving research has successfully assessed psychological processes by assessing EMG (Healey et al., 1999; Fu et al., 2016; cf., Morris et al., 2017; Ma et al., 2018).

EMG in Driving Context

In driving contexts, surface EMG has been utilized to study psychological and physiological stress (Jonsson and Jonsson, 1975; Wikström, 1993; Balasubramanian and Adalarasu, 2007; Ahlström et al., 2018). Stress and fatigue have been studied by recording electrical activity from relevant muscles. For instance, variations in the trapezius muscle (a major back muscle that extends from the neck to shoulder blades and lower spine) and deltoid (triangular muscle located on uppermost part of an arm and the top of shoulder) are influenced by mental stress (Wikström, 1993; Balasubramanian and Adalarasu, 2007; Hirao et al., 2007; Wijsman et al., 2013; Luijcks et al., 2014; cf., Morris et al., 2017). A recent study (Lee et al., 2017a) recorded trapezius muscle activity to detect stress in a driving simulator under relaxed and stressed conditions. A continuous increase over time in muscular tension was associated with greater stress experienced due to driving task (Lee et al., 2017a). Muscular tension can thus be a useful metric of stress level that can be utilized in driving research.

It is worth noting that muscular fatigue and discomfort are not isolated issues (Leinonen et al., 2005) and they cause psychological distress and disrupt cognitive performance while driving. Muscle fatigue while driving has been studied by examining changes in muscular tension in shoulder and neck muscles (Sheridan et al., 1991; Wikström, 1993; Balasubramanian and Adalarasu, 2007; Hirao et al., 2007). Compared to the beginning of the drive, continuous driving can lead to reduced back muscles (e.g., trapezius and deltoid) activity and fatigue. Muscular fatigue (measured by EMG of back muscles) is associated with decreases in power of EMG activity-related frequency band (Hostens and Ramon, 2005; Balasubramanian and Adalarasu, 2007; Hirao et al., 2007). Surface EMG is a helpful way of identifying discomfort in fatigued and weak muscles and targeting rehabilitation for skeletomuscular problems specially in professional or long-distance drivers (Balasubramanian and Adalarasu, 2007). A recent study (Artanto et al., 2017) has also used a low-cost EMG system to detect drowsiness. An EMG sensor attached to muscles around eyelid region captured the duration of eyelid closure as an indicator of drowsiness (Artanto et al., 2017). Another recent study has proposed a system that can detect real-time changes in EMG (Mazzetta et al., 2018). Further research is needed to validate EMG's applicability in real-world settings.

Practical Considerations

EMG measurement enable recording continuous data from the specific muscle of interest without obstructing the driving task. Such objective information can be helpful in learning about muscular activity (and relevant cognitive states) that may not be necessarily visible to the researchers or under the awareness of the responder. However, it is essential to pay attention to any outliers or irrelevant events that may add noise to the EMG signal and impact signal interpretation. Irrelevant events can include muscular activity due to driving-unrelated (e.g., continuous posture change, scratching skin, or touching the electrodes) and driving-related (e.g., functional steering activity) movement and yet unrelated to the cognitive state (e.g., mental workload) of the driver (Mehler et al., 2009). In real-world settings, it can be tedious to tease apart muscular activity due to other confounding reasons from activity relevant to changes in cognitive states. Furthermore, the task under investigation is also of importance. For instance, a study that compared muscular tension while driving car autonomously vs. manually found no differences in EMG signals, but significant differences were found for SCL (Morris et al., 2017). This suggests that for some tasks the muscular activity may not significantly differ, but may still be psychologically different in other modalities. This also highlights the importance of multiple measures.

Thermal Imaging

Thermal Imaging Quantification

The measurement of changes in skin temperature is a useful technique to detect and track attributes of a responder, such as body posture and emotional expression (Gade and Moeslund, 2014; Rai et al., 2017). A special merit of this technology is that it enables sensing the real-time state of motorists non-invasively without disrupting driving related tasks. In addition, unlike RGB cameras, thermal cameras do not depend on an external illumination (Gade and Moeslund, 2014; Rai et al., 2017). Objects that emit radiations in the mid-to-long wavelength infrared spectrum (3–14 μm), such as the human body (but not inanimate objects) can be detected via thermal imaging (Gade and Moeslund, 2014; Rai et al., 2017). Changes in temperature distribution, as captured by the thermal cameras, are utilized to make meaningful interpretations. For instance, facial thermography can be used to capture the heat distribution in facial locations known to vary with sympathetic activity as a metric of the varying psychological phenomena. Most commonly investigated facial locations include the forehead and nasal temperature changes.

Sympathetic autonomous nervous system activation may lead to constrictions of blood vessels, thereby decreasing temperature in extremities, such as the nose (Or and Duffy, 2007; Gade and Moeslund, 2014). For example, mental workload changes lead to temperature variations in the forehead, nose, cheeks, and chin regions (Stemberger et al., 2010; Marinescu et al., 2018). A recent study examined the validity and sensitivity of thermal imaging in assessing variation in cognitive load (Abdelrahman et al., 2017). Increased cognitive task difficulty led to significant increases in the forehead temperature and decreases in nose temperature (Abdelrahman et al., 2017). The largest effect sizes were found when the difference in forehead and nose temperature was estimated. Higher task difficulty led to an increase in forehead and nose temperature differences (Abdelrahman et al., 2017). Additional work has also examined real-time sensitivity of thermal imaging and found that specialized thermal cameras can detect changes in cognitive load with a latency of 0.7 s post eliciting event (Abdelrahman et al., 2017). This finding suggests that this methodology has a high relevance for real-time assessments of cognitive load in applied settings like driving.

Thermography in Driving Context

In driving contexts, facial thermography was found to be useful in assessing over-arousal constructs such as mental workload (Or and Duffy, 2007; Murai et al., 2008). Performing a secondary workload task (mental arithmetic) while driving in a simulator as well as an on-road car led to a decrease in nasal temperature with stable forehead temperatures (Or and Duffy, 2007). Drop in nasal temperature also correlated with self-reported workload (Or and Duffy, 2007). Another study found increases in the difference between nose and forehead temperature increased with mental workload (Kajiwara, 2014). Participants' nasal temperature varied as a function of mental workload in simulated driving (Kajiwara, 2014). Workload variation indexed by changes in nasal temperature were also reported during ship navigation using a simulator (Murai et al., 2008), highlighting its utility in applied settings.

Furthermore, facial thermography can be useful to examine and infer heat distribution in faces during emotional states. This method could be promising and may provide a non-invasive approach to capture emotional states because current methods of emotion recognition using facial features detection software have limitations. One study used an infrared thermal camera to non-invasively detect face regions and recognize emotional states of motorists (Kolli et al., 2011). This study suggests that thermography can improve face detection algorithm for in-vehicle settings thereby facilitating ADAS.

In another line of work (Cheng et al., 2007), a combination of thermal infrared and color cameras have shown to be effective in sensing body movements in real-time on-road driving. Similarly, infrared streaming has been used to develop posture and occupancy sensory systems (Kato et al., 2004; Trivedi et al., 2004). Another recent study reported successful use of near-infrared light and thermal camera sensors to identify aggressive driving behavior (Lee et al., 2018) and were able to categorize aggressive driving from relaxed driving. The above studies suggest that thermography has the potential to be a useful non-invasive technique that can be validated to capture cognition-relevant states and improve traffic safety.

Practical Considerations

Thermal cameras are used in numerous industrial, agricultural, and military settings (Gade and Moeslund, 2014). They can be extremely useful in vehicular technology because they are non-contact sensors and can work regardless of external illumination. Nevertheless, further testing is needed to better understand how this technology would improve our

understanding of cognitive states in traffic safety. Further systematic investigation and replication of thermography as a function of cognitive workload, stress, and drowsiness after controlling for confounding factors, such as environmental factors (e.g., weather conditions and air conditioning), are needed to be able to make confident assessments of cognitive states. The results so far look promising.

Pupillometry

Pupil Quantification

Pupillometry is the measurement of pupil size and reactivity. Modern pupillometry is measured via optical eye-trackers that use some combination of monitoring infrared light reflections from the cornea, the back of the lens, and the pupil, as well as absorption of light by the pupil (e.g., dark-pupil tracking). Most modern eye-tracking devices can monitor pupil location (and eye-fixation location) with very high resolution (>1,000 Hz) non-invasively and at a substantial distance from a participant. Thus, measurement can occur in highly ecologically valid environments, without participants having to make any overt responses. Since the 1960's it has been shown that pupil dilation changes as a result of mental activity—for example, increases in arousal and cognitive workload (e.g., Hess and Polt, 1964). In a classic study demonstrating the sensitivity of pupillometry to cognitive demands, Kahneman and Beatty (1966) showed that pupil dilation increases parametrically with an increasing number of words to recall in a simple word list memory task. Moreover, they showed that this increase in workload persists over a maintenance interval, and reduces parametrically as each word is retrieved (and released) from memory. These findings, along with a number of other demonstrations of pupillary sensitivity to cognitive workload, for example in math problem solving (Sirois and Brisson, 2014), working memory and individual differences in intelligence (Tsukahara et al., 2016), aging and verbal memory load (Piquado et al., 2010), has led to wide interest in this measure as a physiological marker of arousal and cognitive effort.

Janisse (1977) remarked that the eye is the only "visible part of the brain." Indeed, detailed models of the neurophysiology of pupillomotor functioning are developed and growing, including an understanding of the innervation of the sphincter and dilator muscles by the autonomic nervous system (Miller et al., 2005), as well as the neuromodulatory relationship between pupil dilation, activity in the locus-coeruleus (LC; a neuromodulatory nucleus in the dorsal pons of the brainstem strongly linked to phasic and tonic arousal, cognitive control, and monitoring functions), and norepinephrine (Gilzenrat, 2006). For instance, a high correlation (0.6) between spike frequency and pupil diameter has been found, whereby large pupil diameter equates to high LC activity (Rajkowski et al., 1994). Demberg (2013) have also recently reported changes in pupillometry due to linguistically induced cognitive load (e.g., comprehending syntactically demanding sentences). Other recent work has also examined user state related changes in pupil diameter in lab-settings such as variations in valence and arousal (Kassem et al., 2017) and interest in real-time (Jacob et al., 2018).

Pupillometry in Driving Context

Eye-tracking has been used extensively in studying visual perception and attention in driving contexts, however the unique use of pupillometry as an index of real-time physiological indicator of cognitive workload is only lately growing in popularity (Schwalm et al., 2008). For example, Cegovnik et al. (2018) recently validated a low-cost eye-tracker and showed that pupil dilation increases with increments in cognitive load due to a secondary memory task (n-back) (see also Recarte and Nunes, 2000 for similar results). Pupillometry has also been adopted in driving research while motorists drove in a simulated driving context. Pupil diameter was found to reliably increase with increases in cognitive load (Palinko et al., 2010; Faure et al., 2016). Other work has use machine learning algorithms to detect cognitive load while driving from pupillometry data (Yoshida et al., 2014). A recent study found that during simulated driving, pupil dilation could detect increases in cognitive load imposed by a secondary task within a lag of 1 s (Prabhakar et al., 2018). This suggests that pupillometry could be used as a near-real time index of cognitive load.

Pupillometry has also been used to differentiate between alertness and drowsiness (Soares et al., 2013). Alertness is associated with increased mean pupil diameter and decreases in standard deviation (i.e., stable), whereas drowsiness is associated with decreases in diameter, but increases in standard deviation (i.e., fluctuations) in pupil diameter (Morad et al., 2000; Wilhelm et al., 2009). Fluctuations in pupil size have been proposed to be a reliable index of drowsiness-related impairment while driving (Maccora et al., 2018). Pupil dilation was also found sensitive to fatigue levels while driving with a decrease in fatigue being associated with an increase in pupil diameter (Schmidt et al., 2017). Although early, these findings, along with others (for a recent review see Marquart et al., 2015; Maccora et al., 2018) suggest that pupillometry is an efficient, ecologically valid, and low-cost physiological reporter variable for indexing cognitive states in driving in highly-controlled environments like realistic driving simulators.

Practical Considerations

In lab settings, pupil diameter was found to be a reliable, non-invasive, and real-time measure of workload (Marinescu et al., 2018). However, in on-road settings, it is quite challenging to capture interpretable pupil information due to large variations in luminance that are hard to control across conditions and participants. Indeed, photopupillary reflex is massive in magnitude relative to changes in pupil size related to cognitive and attentional factors. As such, if there are considerable changes in lighting conditions (e.g., sunny vs. cloudy days), this can create considerable noise in the pupillary signal. Moreover, if specific conditions of interest are confounded with respect to overall luminance (e.g., driving during the day vs. driving at night), this overall pupillary light reflex-related shift should be taken into consideration. Furthermore, if investigating event-related pupillary responses in driving, one should be careful to determine that differences in pupil dilation are not only due to differences in visual stimulation (e.g., presenting a luminant STOP sign). Modeling techniques have also developed methods

to infer cognitive workload after accounting for some variations in lighting conditions (Pfleging et al., 2016; Reilly et al., 2018).

Marshall (2002) have developed a signal processing method for extracting high-frequency changes in pupil dilation that they argue is uniquely related to cognitive components (*Index of Cognitive Activity* or ICA). However, this method is a commercially available "black box" system, and should be interpreted with caution given that the exact algorithm used to calculate ICA from raw pupillometry is not open source. Other work has estimated an *Index of Pupillary Activity* (IPA) inspired by ICA, that uses wavelet-based algorithms to decompose pupil data (Duchowski et al., 2018). IPA was found to differentiate between low vs. high mental workload (Duchowski et al., 2018). Another important feature to consider is that measurement of pupil dilation is affected by eye-movements and relative gaze position (e.g., Gagl et al., 2011). When gaze position changes from central to peripheral locations, the recorded pupil shifts from a circular to an elliptical shape from the point of view of fixed camera location. This change in the recorded geometry of the pupil is accompanied by changes in overall pupil size, irrespective of actual changes in dilation or constriction. Gagl et al. (2011) have developed methods for the measurement and removal of such systematic influences. Nevertheless, researchers should be careful to measure gaze position and to design studies such that likely visual target locations are not confounded across conditions of interest.

CHALLENGES AND RECOMMENDATIONS

Psychophysiological research has made tremendous progress in developing methods to quantify cognitive processes. Most of this research has been conducted in carefully controlled environments to be able to interpret with certainty what changes in a physiological signal may imply about the psychological phenomena under investigation. Physiological signals are valuable to understand how people interact in real-world contexts. Driving research is an excellent application of psychophysiological methods to understand and interpret how people interact with automation in natural settings, which in turn can inform intelligent systems to improve driving performance and safety. As evidenced by much of the growing research base discussed above, psychophysiological measures can be successfully adopted to meet these goals. At the same time, lack of adherence to research protocols and guidelines can seriously jeopardize meaningful use of these methodologies. Here we highlight a few general challenges and recommendations that cut across all psychophysiological measures in driving research when collecting data from real-world driving settings—which are less predictable than lab settings— to improve data-quality and aid in effective interpretation.

Valid and Reliable Quantification of Construct

Depending upon the task and setting (lab-based simulator or field study), some physiological measures will be more suitable and feasible than others. For example, in a simulator with very controlled body movement, continuous blood pressure using the volume clamp method can be collected. However, while on-road, this equipment may compromise drivers' safety and thus is not feasible. Other measures like ECG and thermal cameras are highly mobile and feasible. Careful observations can allow interpretation of cognitive processes while driving. One important concern is the possibility of misinterpreting the relationship between physiological signals and cognitive processes (Cacioppo and Tassinary, 1990; Cacioppo et al., 2007). Often, physiological measures (such as HR, EDA, EMG) are impacted by multiple processes, such as drowsiness, stress, and workload, which can lead to interpretive caveats. Systematic variations in different experimental conditions can help tease apart the underlying mechanism causing autonomic activations to be able to draw clear inferences. However, in an applied setting like driving a car in unpredictable traffic, control over the experimental task is largely out of the control of the researcher. Confirmatory independent measures are important to validate the construct of interest in the study. Similarly, it is helpful to ensure that the construct of interest reliably varies across conditions and that the experimental manipulation was effective.

Individual Differences

A combination of factors may influence physiological signals, including trait-level variables such as demographic factors (age, gender), task experience (professional, experienced, inexperienced), anxiety, and certain health conditions and medications (e.g., cardiovascular health). State-level variations such as stress-levels unrelated to task, caffeine intake (which may change autonomic activity), and engagement/motivation and frustration during the task can also interact with individual differences in ways that may not be readily apparent. Combining data from participants after considering such trait- and state-level variables can help in proper interpretation of study findings.

On a related note, a critical challenge in multi-modal recordings is that individuals may be highly reactive as assessed by one measure but not necessarily, according to another. There is considerable variability across individuals in how closely physiological, behavioral, and subjective measures covary over time with one another (Lohani et al., 2018). Furthermore, it is possible that only some individuals may be sensitive to the experimental manipulation (Drews et al., 2012). Such individual differences may lead to variations in psychophysiological assessments and may also explain to some extent lack of significant differences across experimental conditions. Many, if not all, of these measures are currently utilized within paradigms where we are studying relative changes in the outcome across conditions (e.g., P3b amplitude is a difference wave, HRV% change, %signal change in BOLD response, etc.), for which these measures do not have currently well-understood absolute thresholds for making strong absolute judgements. While there isn't a fixed threshold for physiological measures that can be used across individuals to define high and low arousal levels, relative changes from baseline can be a useful way of assessing variations in arousal levels from optimal levels for the individual. If the system can be calibrated on what is a "normal" range for an

individual, then significant variations from this calibrated range can be a way to detect sub-optimal arousal levels.

Baseline Assessments

Baseline assessments provide insights about the physiological state of the responder when the experimental condition was absent. It also allows to control for physiological activity due to any prior conditions, so that the change in the experimental condition of interest is interpreted relative to the state right before the condition started. A single baseline is generally not enough, especially when there are multiple conditions. It is a good practice to capture as many baseline assessments and as close to the experimental condition as possible. Another alternate design to consider (for measures with high temporal resolution) is an event-related design, where activity is time-locked to specific events of interest. In this design, pre-event activity in the measure is subtracted from the overall physiological time series, resulting in a strong baseline control for each trial (e.g., ERPs).

Sampling Rate, Filtering, and Signal Quality

Nearly all physiological signals discussed above are analog signals, which have to be digitized for further processing. Choice of optimal sampling rate and filtering helps avoid signal distortions (Jennings and Allen, 2016), and as such, knowledge of signal processing characteristics of the target physiological measures is necessary for researchers to effectively use these tools. Optimal sampling rate differs by the physiological signal's frequency characteristics, and poor sampling rate can distort waveform characteristics, and induce artificial oscillatory characteristics that are not part of the true analog signal (i.e., aliasing). For example, for HRV analysis, the recommended sampling rate is at least 250 Hz (Task Force of the European Society of Cardiology, 1996). Some commercial wearables (e.g., fitness-related wrist watch sensors) have sampling rate as low as 60 Hz, which will lead to signal aliasing (Jennings and Allen, 2016) and inaccurate and uninterpretable HRV values. The sampling rate needs to be at least above the Nyquist frequency (2x the sampling rate of the highest frequency), and current standards suggest a sample rate 3–4 times the highest frequency component of physiological signal. Advancements in modern computing allow for research-grade equipment to sample far above Nyquist for most of the measures discussed (>2,000 Hz) during data acquisition. Of course, data can always be down-sampled post data collection. As discussed in sections "Heart Activity Quantification" and "Practical Considerations" on heart activity, quantification using wearables can lead to inaccurate assessments (Laborde et al., 2017) due to poor sampling rates, lagged responses, and noisier signals to name a few, which would lead to inaccurate interpretations.

Filters are helpful in getting rid of artifacts and noise not relevant for the physiological signal being processed. For instance, muscle and electrical noise (around 60 Hz) are not meaningful while interpreting EEG and ERP data, and thus data outside the range of interest (typically not higher than 40–50 Hz) can be bandpass filtered. However, if EMG activity, which has a much higher frequency content, is of interest, then bandpass filtering with allow low-pass cutoff at 500 Hz and high-pass cutoff at 20 Hz, is often suitable (van Boxtel, 2001). Visual inspection pre- and post-filtering process can help determine how filtering is affecting a signal. Note that all filters distort the waveform and spectral characteristics, so unnecessary filtering should be avoided and researchers should take care to understand exactly how filters are impacting their data in time and frequency domains.

For each psychophysiological measure discussed, researchers have a growing number of indices that can be examined (for example, for HRV, time-based, frequency-based, and non-linear measures can be derived). Choice of metrics should be carefully evaluated, as some metrics may be more suitable to meet the goals of the study, while others may not be suitable. For instance, some metrics require minimum duration of data and falling short of such requirements will lead to misrepresentative findings (e.g., standard deviation of R-R heart beats or SDRR is considered more accurate when calculated over 24 h vs. 5 min or shorter intervals; Shaffer and Ginsberg, 2017). Such choices should be made a priori, based on the research question of interest and links between a measure and its purported psychological interpretation based on prior research. Such flexibility in multi-modal recording comes at the cost of an increasing number of "experimenter degrees of freedom," that can lead to inflated Type-I error rates, if a consistent analysis pipeline is not followed. It is also important to use comparable durations of physiological signals across conditions and participants for appropriate interpretation. Finally, great attention to accurate event markers is critical for valid interpretation within and across participants in event-related designs. This can be an issue when using commercial products that are not designed for research purposes.

Innovation

A limitation of most current psychophysiological research-grade measures is the need for using contact sensors (placed on skin). Non-contact sensors are beginning to be tested in applied settings, which can make physiological data collection even less invasive. For instance, ECG data can be derived from high-quality RGB cameras, or sensors could be placed on the steering wheel and driving seats (but should meet the recommended requirements). While these can potentially be a great approach to counter the limitations of contact sensors, caution is advised while considering them because new limitations or inaccuracies in assessment are possible and further research and testing is required to adopt them in research. Commercial products may not meet the requirements recommended by the scientific community, which can lead to poor data quality and invalid interpretations. For example, smartphone camera-based PPG sensing estimates have poor sampling rate and can lead to inaccurate assessments (Laborde et al., 2017). It is essential to ensure that the guidelines for measures are met before investing time and resources to avoid technical issues in data collection and interpretation. For instance, as discussed earlier, it is critical to collect physiological data with recommended frequency sampling to avoid aliasing (Jennings and Allen, 2016). Only equipment that have been or can be validated against research-grade devices should be adopted for research purposes.

Classification

Reliable and valid assessment of cognitive states is the groundwork to develop inputs to advance state detection-workload managers and "aware" systems. For instance, a recent study reported a reliable method to elicit stress in naturalistic driving scenarios (Baltodano et al., 2018). Given that one measure may not be enough to reliably measure subtle changes in cognitive state, a multi-method approach is critical to capture state-level variations that may not be apparent through a single measure alone. Research has shown that multi-modal approaches provide a reliable (Schmidt et al., 2011; Borghini et al., 2014; Chen et al., 2017) way to sense and assess cognitive states of motorists in real-world settings. Notably, due to the dynamic nature of the physiological signals, conventional linear approaches are not always appropriate in modeling and predicting cognitive state (Chen et al., 2015). The discussed physiological signals are often non-stationary overall but for the briefest periods of time. As such, innovative methods of combining temporal and spectral resolution (time-frequency analysis) have been developed in some domains (e.g., EEG), but their application to other physiological signals is only in its infancy.

Once data have been processed to remove artifacts or irrelevant noise, machine learning techniques could be trained on these data to identify "risky" sub-optimal levels of cognitive states, such as low-arousal states of drowsiness and fatigue associated with unsafe driving performance. During the training phase, multimodal features extracted from physiological training data could be used to train models to classify observations into high-arousal states (e.g., due to high stress and workload), optimal-arousal state, vs. low-arousal state (e.g., due to drowsiness and fatigue). During the test phase, the fully-specified machine learning algorithm can be tested in terms of its capacity to accurately classify observations into respective arousal states. Indeed, cognitive state detection based on multimodal feature analysis and classifiers have been also used to detect stress (Yang et al., 2016; Chen et al., 2017; Lee et al., 2017b), alertness and drowsiness (Forsman et al., 2013; Correa et al., 2014; Chen et al., 2015; Wang and Chuan, 2016), fatigue (Fu and Wang, 2014; Wang, 2015; Fu et al., 2016; Li et al., 2017; Wang et al., 2017), and workload (Borghini et al., 2014; Yang et al., 2016) in real-time. Such studies have integrated data from more than one measure by conducting multi-modal analysis to extract the relevant features to capture the psychological phenomena at hand. A comparison of multiple classifiers to train & optimize machine learning algorithms can help determine the best fitting model to represent changes in cognitive states that can explain driving performance (Nadeau and Bengio, 2000; Fairclough et al., 2015; Balters and Steinert, 2017; Tran et al., 2017). Thus, utilizing multi-modal physiological signals, models could be trained to learn and predict motorists' sub-optimal cognitive states associated with unsafe-driving behavior.

The optimized machine learning algorithms could accordingly inform advanced state detection managers to trigger warnings or otherwise intervene when sub-optimal cognitive states associated with risky driving behavior are detected (Aidman et al., 2015). The ability to predict unsafe levels of physiological arousal will enable targeted augmentation to modify motorists' cognitive state to promote safer driving behavior (Schmidt and Bullinger, 2017; Schmidt et al., 2017; Aricò et al., 2018). For instance, countermeasures to augment cognitive states, such as thermal stimulation (Schmidt and Bullinger, 2017; Schmidt et al., 2017) and warning signs or verbal communication (Schmidt et al., 2011; Aidman et al., 2015) can be used by an automated system to modify drivers' cognitive state. This may especially benefit vulnerable groups such as inexperienced drivers (Noordzij et al., 2017; Yan et al., 2017) and older (Costa et al., 2017) drivers who may be more susceptible to cognitive overload. Furthermore, a person-centered approach can account for individual differences, such as the role of age, driving profile, trust, and reliance on automation. For instance, a recent study used discriminant analysis to account for motorists' driving-styles and individual difference factors (e.g., gender, age, anxiety, anger) and also identify motorists' EEG and EDA response features to classify motorists' safe vs. risky driving tendencies (Liang and Lin, 2018). This study shows that individual differences can explain variations in driving performance and a customized approach may also help improve model prediction over time by accounting for motorists' characteristics and preferences. For example, the low, normal, and high physiological arousal ranges will vary depending on attributes such as anxious, risky, and distress reduction driving styles of an individual (Liang and Lin, 2018) and prediction of cognitive state-level variations may be more accurate when predictions account for such individual-level variations. Thus, a person-centered approach will improve reliable predictions of cognitive states in real-world contexts by intelligent driving systems.

RESEARCH APPLICABILITY IN REAL-WORLD SETTINGS

As the reviewed literature in section, "Psychophysiological Measures to Assess Cognitive States" suggests, many interrelated states could lead to a similar pattern of findings on a physiological measure (e.g., mental fatigue, drowsiness, lower vigilance, and mind wandering are all sensitive to similar EEG/ERP indices). After considering the overlap across findings from interrelated constructs, in **Table 1** we have summarized the expected pattern that each physiological measure will have during a low vs. high arousal state in an applied driving context. There are a few points to consider. First, changes in several related cognitive states can lead to similar changes in arousal. For example, increases in driver workload, stress, or vigilance may occur under different contexts, but may similarly lead to heightened arousal. Second, even though arousal is continuous, we chose to classify driver states into categories of low and high arousal because both extremes are sub-optimal for driving performance. Third, cognitive states are complex and change across time. For instance, in the current review, we have placed mind wandering in a low-arousal state based on similar patterns of findings as drowsiness. However, mind wandering is a convenient short-hand for a

more complex constellation of non-externally directed cognitive states (see Smallwood and Schooler, 2006 for a review) and depending on the context, such mind-wandering states can yield states of heightened-arousal as well. Similarly, fatigue can be categorized as high-arousal due to prolonged cognitive overload or it can be passive, because of underload due to monotonous driving conditions, for example (Saxby et al., 2008; Matthews et al., 2019). With further empirical evidence in naturalistic environments, a better characterization of complex cognitive states could be developed.

It is still an open question if interrelated cognitive states could be successfully differentiated from other similar states in naturalistic environments (see Cacioppo et al., 2007 for challenges with psychological inference). However, physiological measures could be used to assess sub-optimal levels of general arousal in real-world settings and intelligent systems can use this information to trigger augmentation strategies even if we cannot fully differentiate between specific cognitive states besides along their arousal axis. We have reviewed how physiological responses across multiple measures can provide a rich array of response data relevant to domains that are of interest to driving researchers (e.g., attention, fatigue, workload, etc.). These measures provide unique information and unique sensitivity to experimental manipulations beyond behavioral responses alone. Thus, their current and future utility in real-world driving research is important. This does not mean that measuring one or even a large number of these measures alone will provide us with a *direct* interpretation of a covert state (e.g., becoming increasingly frustrated about an aggressive driver behind you). Before the state of the research matures to be able to address such a lofty goal as predicting specific cognitive states (Yarkoni and Westfall, 2017), we first need careful on-road experimental work to understand the sensitivity and specificity of these measures to specific changes in driver-relevant states in observational and experimental research in real-world settings. Thus, the focus of the current review is not to claim that measurement of multiple physiological measures in real-world driving could accurately predict motorists' specific cognitive state. Rather, our goal is to summarize the feasibility of each of these measures for integrating high-quality psychophysiological methodology into real-world driving research. **Table 1** presents the current working predictions that are expected based on the available literature, but more work is needed to be able to use physiological signals to infer psychological processes. The current review represents a summary of initial steps in that direction.

In **Table 2**, we have summarized the research applicability of the reviewed psychophysiological measures. Although all of these measures can provide valuable insights in the controlled settings of a lab, some measures are more feasible to use and interpret than others in real-world driving contexts. A few factors that may play a role in determining the practical use of physiological measures in applied settings are: the degree of coupling between the measure and subtle changes in cognitive states, temporal resolution, psychometric reliability, ease of data collection (e.g., setup time), sensitivity to artifacts, and the degree of invasiveness and disruption to normal driving. After considering the available evidence, we have categorized each measure's real-world research applicability into low, medium, or high levels. Moreover, certain measures may be better candidates than others for a near real-time assessment in applied settings. We review the real-world applicability and feasibility of each of the measures in **Table 2**.

Some promising work suggests that cardiovascular measures may be robust in detecting near real-time changes across multiple domains. Studies have shown that cardiovascular data can reliably detect changes in workload (Mehler et al., 2009, 2012; Lenneman and Backs, 2010; Stuiver et al., 2014), fatigue (Patel et al., 2011; Matthews et al., 2019), and drowsiness (Vicente et al., 2016; Kurosawa et al., 2017). Like any physiological signal, cardiovascular data is susceptible to artifacts that could otherwise lead to inaccurate estimations. However, recent analytical advances have led to an improved use in real-world settings even in the presence of substantial recording artifact. For instance, an analysis approach using short segments of cardiovascular data (e.g., a moving window of 30 s; Stuiver et al., 2012) can be used to detect workload demands during driving (Stuiver et al., 2014). Use of smaller temporal windows of data allow for an investigation of the short-term effects of cognitive state without being overly susceptible to artifacts. Recent work has shown that frequency analysis techniques on ECG data can also be utilized to detect early onset of fatigue (Matthews et al., 2019). While the limitations of PPG discussed earlier still apply, recent preliminary work using near-infrared illumination PPG (which overcomes confounds of illumination and motion-related inaccuracies) while driving seems a promising direction for future practical applications (Nowara et al., 2018). Another recent work has developed a noise-resistant algorithm specifically designed to analyze PPG waveforms (van Gent et al., 2018), which can provide researchers an open-source and validated heart rate analysis software to overcome some existing limitations of PPG data processing, making it more feasible for applied driving research.

EDA has been found to be a robust measure of sympathetic arousal in driving contexts in real-world settings (Mehler et al., 2012; Schneegass et al., 2013; Ruscio et al., 2017). EDA is also easy to set up and collect from a motorist without obstructing the driving process. Even though it has a slower response time and provides only a broad sense of arousal (a combination of workload, stress, fatigue, etc.), EDA in an applied uncontrolled environment can estimate relative changes and periods of stability in sympathetic activity of a motorist with an upper temporal resolution of approximately 3–5 s. For example, recent work found EDA to be suitable in capturing stress-level variations in a real-time unconstrained setting (ElKomy et al., 2017). Feature extraction and pattern recognition algorithms have also shown reasonable success recently in detecting changes in cognitive states (Chen et al., 2017; Liu and Du, 2018). Moreover, adaptive filters have been successfully used to remove motion-related artifacts for automatic and accurate detection (up to 95% sensitivity) of state-level variations in cognition (Affanni et al., 2018). Such recent processing and analytic advances with EDA data has shown its high relevance in applied intelligent automation. For example, a development approach proposed for monitoring driver's fatigue levels and functional state utilizes automated analysis of EDA indices in their detection module

to improve intelligent vehicular systems (Liu and Du, 2018; Savchenko and Poddubko, 2018).

EEG, a direct measure of brain's electrical activity, can provide robust measures of cognitive state variations while driving, including levels of drowsiness (Liang et al., 2006; Wei et al., 2018), fatigue (Liu et al., 2015; Fu et al., 2016; Hung et al., 2017), and workload (Dasari et al., 2017; Zander et al., 2017). EEG has high temporal resolution and is a direct measure of brain activity. However, data collection (e.g., longer setup time) and processing in real-world setting (e.g., movement artifacts) can be quite challenging to implement into a real-world driving research protocol (Popescu et al., 2008). At the same time, there have been innovative technological and analytical developments in EEG acquisition. For instance, efforts in brain computer interface applications have utilized a single electrode to classify relaxed vs. cognitive workload phases (Shirazi et al., 2014) and monitor fatigue levels (Morales et al., 2017). Recent work extracted features from a 6-channel EEG dataset to classify mental tasks with up to 83% accuracy rate (Neshov et al., 2018). Other recent work has reported detection algorithms that can be used to accurately classify fatigue (Li et al., 2017; Gao et al., 2018). In other work, a novel approach to detect drowsiness has been proposed which reduces calibration time for a new user by 90% using a hierarchical clustering method, which accounts for inter- and intra-subject variability (Wei et al., 2018). Automatic drowsiness detection algorithms based on only a single target channel can allow real-time neural assessments of cognitive states (Belakhdar et al., 2018). With increasing advancements in sensor development and data processing, we hold an optimistic view of adopting EEG-based measures in driving research, albeit after considerable validation (Kosiachenko and Si, 2017; Krol et al., 2017; Zander et al., 2017; Byrom et al., 2018). Recent work has also shown the applicability of specific ERP components (such as the P300), some of which show good psychometric properties (e.g., Cassidy et al., 2012), and can be adopted to brain-computer interfaces (Piña-Ramírez et al., 2018). Future work and reliable replication of studies are required to ensure EEG and ERPs could be assimilated in human-machine automation interface.

Traditional fNIRS has lower temporal resolution and may additionally be difficult to collect in applied settings. However, recently, mobile-friendly systems have been developed and used in applied domains (von Lühmann et al., 2015) including exercise physiology (Byun et al., 2014), clinical monitoring (Kassab et al., 2018), and infant developmental research (Quaresima et al., 2012). Importantly, these advancements mean that fNIRS measurements can be performed in naturalistic environments without considerable restraint. As the development of ultra-portable systems grows (e.g., battery powered mobile systems, McKendrick et al., 2016), fNIRS will likely form a novel complement to the many other physiological measures discussed here, in part because of its unique capability to image neural hemodynamics and reveal changes in brain activity with improved spatial resolution compared to other portable and non-invasive neurophysiological methods (e.g., EEG; Ahn and Jun, 2017). For instance, a recent study adopted a wearable fNIRS system (with sensors placed on a baseball cap making it less intrusive) to measure cognitive distraction while driving (Le et al., 2018). Thus, while these methods are still in their infancy compared to many of the other methods discussed here, the ability to reveal neural mechanisms of cognitive states in real-world domains such as driving is promising.

Similar to fNIRS, thermal imaging also shows some early promise. It is a non-contact technology that has high relevance in applied settings, including driving (Lee et al., 2018). For example, recent work has shown the validity of thermal imaging in indexing cognitive load. In these studies, changes in nasal and forehead temperatures were observed as a function of task difficulty in a non-driving context (Abdelrahman et al., 2017; Marinescu et al., 2018). However, research in real-world settings is currently limited. Existing preliminary work has focused primarily on understanding the sensitivity of this measure in well-controlled environments. Future work will help qualify the utility and validity of thermal imaging in real-world conditions.

On the other hand, several measures, despite clear utility in a lab environment, may be currently of less use in real-world settings. For example, pupillometry in well-controlled lab settings can provide helpful information in interpreting user state (e.g., Pfleging et al., 2016; Cegovnik et al., 2018). Moreover, with the development of desktop-mounted eye trackers, pupil dilation and constriction can be measured non-invasively and remotely with high spatial and temporal resolution. In lab settings, where features such as luminance can be controlled and measured, recent work has shown success in using pupillometry to examine mental workload in an unconstrained setting (e.g., Lego construction; Bækgaard et al., 2019). In driving, some researchers have suggested that pupil-based measurements are highly relevant for assessment of drowsiness (Maccora et al., 2018). However, detection of pupil diameter in real-world settings with rapidly changing and uncontrollable variations in luminance is a critical confounding factor in the utility of pupillometry in driving (Kassem et al., 2017).

Similarly, EMG can be utilized in lab settings to understand psychological processes. For example, EMG in combination with other psychophysiological measures was recently utilized in detecting fatigue in drivers (Fu et al., 2016; Ma et al., 2018). Preliminary research has also proposed the use of EMG to detect drowsiness (Artanto et al., 2017) and real-time monitoring of muscle activity (Mazzetta et al., 2018). However, in applied settings such as driving, EMG may have only low utility, in part because the necessary motor activity needed to engage in the task (e.g., turning the steering wheel and actuation of break) can cause uncontrolled changes in muscle activity that can be confounded with the psychological variance in EMG, which is an order of magnitude smaller than these artifacts.

At the same time, ongoing methodological developments are resulting in more efficient systems, improved signal-to-noise ratio, and improved signal-processing methods, all of which culminate in rapidly improving the reliability and validity of acquisition across these multiple methodologies. Some attempts to assess cognitive states using multiple methods have been integrated in non-driving domains (ElKomy et al., 2017; Ko et al., 2017; Moghaddam and Lowe, 2019) and multi-method work in real-world driving contexts are already underway (Fu et al., 2016; Brouwer et al., 2017; Zander et al., 2017; Aricò et al., 2018;

Belakhdar et al., 2018; Haouij et al., 2018; Paredes et al., 2018; Rastgoo et al., 2018).

Taken together, we have reviewed a growing body of empirical evidence suggesting that physiological measures can be used to sense and assess changes in the cognitive states of motorists during real-world driving. Through this selective review, we believe that the strengths and limitations of adopting physiological measures in driving can clearly extend to other domains such as the use of aircraft, trains, and ships. Furthermore, we see growing promise for the application of covert monitoring methods like those reviewed above with the increasing rise in semi-automated technology, where motorists will become less directly involved in the driving process. As such, the development of intelligent driving assistance systems will need to utilize non-behavior-based measures to index covert cognitive states of a motorist in the absence of any overt behavior. The physiological measures reviewed above have the potential to detect sub-optimal arousal levels associated with risky driving behavior and inform state detection-workload managers and "aware" systems to trigger warnings or intervene, resulting in a closed-loop system in the absence of any overt-driving behaviors. Before we reach such a future however, the field needs to adopt rigorous standards for the use of psychophysiological measurement in real-world settings. We hope to see a future of increased collaboration and integration of basic psychophysiology, human factors, and traffic safety research. Such integration is necessary to advance the development of effective human-machine driving interfaces and driver support systems, with the ultimate goal of improving traffic safety.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

Support for this paper was provided by a grant from AAA Foundation for Traffic Safety.

REFERENCES

AAA Foundation for Traffic Safety (2018). *2017 Traffic Safety Culture Index*. Washington, DC: AAA Foundation for Traffic Safety.

Abdelrahman, Y., Velloso, E., Dingler, T., Schmidt, A., and Vetere, F. (2017). "Cognitive heat: exploring the usage of thermal imaging to unobtrusively estimate cognitive load," in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies,* (New York, NY). doi: 10.1145/3130898

Abe, E., Fujiwara, K., Hiraoka, T., Yamakawa, T., and Kano, M. (2016). Development of drowsiness detection method by integrating heart rate variability analysis and multivariate statistical process control. *SICE J. Control Meas. Syst. Integr.* 9, 10–17. doi: 10.9746/jcmsi.9.10

Affanni, A., Bernardini, R., Piras, A., Rinaldo, R., and Zontone, P. (2018). Driver's stress detection using skin potential response signals. *Measurement* 122, 264–274. doi: 10.1016/j.measurement.2018.03.040

Ahlström, C., Gink Lövgren, M., Nilsson, M., Dukic Willstrand, T., and Anund, A. (2018). The effect of an active steering system on city bus drivers muscle activity. *Int. J. Occup. Saf. Ergonom.* 2018, 1–9. doi: 10.1080/10803548.2018.1445465

Ahn, S., and Jun, S. C. (2017). Multi-modal integration of EEG-fNIRS for brain-computer interfaces–current limitations and future directions. *Front. Hum. Neurosci.* 11:503. doi: 10.3389/fnhum.2017.00503

Aidman, E., Chadunow, C., Johnson, K., and Reece, J. (2015). Real-time driver drowsiness feedback improves driver alertness and self-reported driving performance. *Accid. Anal. Prev.* 81, 8–13. doi: 10.1016/j.aap.2015.03.041

Angelidis, A., van der Does, W., Schakel, L., and Putman, P. (2016). Frontal EEG theta/beta ratio as an electrophysiological marker for attentional control and its test-retest reliability. *Biol. Psychol.* 121, 49–52. doi: 10.1016/j.biopsycho.2016.09.008

Aricò, P., Borghini, G., Di Flumeri, G., Sciaraffa, N., and Babiloni, F. (2018). Passive BCI beyond the lab: current trends and future directions. *Physiol. Meas.* 39:08TR02. doi: 10.1088/1361-6579/aad57e

Arnau, S., Möckel, T., Rinkenauer, G., and Wascher, E. (2017). The interconnection of mental fatigue and aging: an EEG study. *Int. J. Psychophysiol.* 117, 17–25. doi: 10.1016/j.ijpsycho.2017.04.003

Artanto, D., Sulistyanto, M. P., Pranowo, I. D., and Pramesta, E. E. (2017). "Drowsiness detection system based on eye-closure using a low-cost EMG and ESP8266," in *Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2017 2nd International Conferences* (Yogyakarta), 235–238. doi: 10.1109/ICITISEE.2017.8285502

Aston-Jones, G., and Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28, 403–450. doi: 10.1146/annurev.neuro.28.061604.135709

Bækgaard, P., Jalaliniya, S., and Hansen, J. P. (2019). Pupillary measurement during an assembly task. *Appl. Ergon.* 75, 99–107. doi: 10.1016/j.apergo.2018.09.004

Balasubramanian, V., and Adalarasu, K. (2007). EMG-based analysis of change in muscle activity during simulated driving. *J. Bodywork Mov. Ther.* 11, 151–158. doi: 10.1016/j.jbmt.2006.12.005

Baldwin, C. L., Roberts, D. M., Barragan, D., Lee, J. D., Lerner, N., and Higgins, J. S. (2017). Detecting and quantifying mind wandering during simulated driving. *Front. Human Neurosci.* 11:406. doi: 10.3389/fnhum.2017.00406

Balters, S., Murnane, E. L., Landay, J. A., and Paredes, P. E. (2018). "Breath booster!: exploring in-car, fast-paced breathing interventions to enhance driver arousal state," in *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare* (New York, NY). doi: 10.1145/3240925.3240939

Balters, S., and Steinert, M. (2017). Capturing emotion reactivity through physiology measurement as a foundation for affective engineering in engineering design science and engineering practices. *J. Intell. Manuf.* 28, 1585–1607. doi: 10.1007/s10845-015-1145-2

Baltodano, S., Garcia-Mancilla, J., and Ju, W. (2018). "Eliciting driver stress using naturalistic driving scenarios on real roads," in *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Toronto, ON). doi: 10.1145/3239060.3239090

Barnard, M. P., and Chapman, P. (2018). The effects of instruction and environmental demand on state anxiety, driving performance and autonomic activity: are ego-threatening manipulations effective? *Transp. Res. Part F Traffic Psychol. Behav.* 55, 123–135. doi: 10.1016/j.trf.2018.02.040

Belakhdar, I., Kaaniche, W., Djemal, R., and Ouni, B. (2018). Single-channel-based automatic drowsiness detection architecture with a reduced number of EEG features. *Microprocess. Microsyt.* 58, 13–23. doi: 10.1016/j.micpro.2018.02.004

Belyusar, D., Mehler, B., Solovey, E., and Reimer, B. (2015). Impact of repeated exposure to a multilevel working memory task on physiological arousal and driving performance. *Transp. Res. Rec. J. Transp. Res. Board* 2518, 46–53. doi: 10.3141/2518-06

Berntson, G. G., Quigley, K. S., and Lozano, D. (2007). Cardiovascular psychophysiology. *Handb. Psychophysiol.* 3, 182–210. doi: 10.1017/CBO9780511546396.008

Billman, G. E. (2013). The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. *Front. Physiol.* 4:26. doi: 10.3389/fphys.2013.00026

Biondi, F. N., Lohani, M., Hopman, R., Mills, S., Cooper, J. M., and Strayer, D. L. (2018). "80 MPH and out-of-the-loop: effects of real-world semi-automated driving on driver workload and arousal," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Sage, CA; Los Angeles, CA: SAGE Publications).

Blöcher, T., Schneider, J., Schinle, M., and Stork, W. (2017). "An online PPGI approach for camera based heart rate monitoring using beat-to-beat detection," in *Sensors Applications Symposium (SAS), 2017 IEEE*. doi: 10.1109/SAS.2017.7894052

Borghini, G., Aricò, P., Di Flumeri, G., Salinari, S., Colosimo, A., Bonelli, S., et al. (2015). "Avionic technology testing by using a cognitive neurometric index: a study with professional helicopter pilots," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE* (Milan). doi: 10.1109/EMBC.2015.7319804

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., and Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* 44, 58–75. doi: 10.1016/j.neubiorev.2012.10.003

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., et al. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology* 49, 1017–1034. doi: 10.1111/j.1469-8986.2012.01384.x

Broadhurst, P. L. (1959). The interaction of task difficulty and motivation: the Yerkes Dodson Law revived. *Acta Psychol.* 16, 321–338. doi: 10.1016/0001-6918(59)90105-2

Brookhuis, K. A., and de Waard, D. (1993). The use of psychophysiology to assess driver status. *Ergonomics* 36, 1099–1110. doi: 10.1080/00140139308967981

Brookhuis, K. A., and de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accid. Anal. Prev.* 42, 898–903. doi: 10.1016/j.aap.2009.06.001

Brouwer, A., Snelting, A., Jaswa, M., Flascher, O., Krol, L., and Zander, T. (2017). "Physiological effects of adaptive cruise control behaviour in real driving," in *Proceedings of the 2017 ACM Workshop on An Application-oriented Approach to BCI out of the Laboratory* (Limassol). doi: 10.1145/3038439.3038441

Byrom, B., McCarthy, M., Schueler, P., and Muehlhausen, W. (2018). Brain monitoring devices in neuroscience clinical research: the potential of remote monitoring using sensors, wearables, and mobile devices. *Clin. Pharmacol. Ther.* 104, 59–71. doi: 10.1002/cpt.1077

Byun, K., Hyodo, K., Suwabe, K., Ochi, G., Sakairi, Y., Kato, M., et al. (2014). Positive effect of acute mild exercise on executive function via arousal-related prefrontal activations: an fNIRS study. *Neuroimage* 98, 336–345. doi: 10.1016/j.neuroimage.2014.04.067

Cacioppo, J. T., and Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *Am. Psychol.* 45:16. doi: 10.1037/0003-066X.45.1.16

Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G. (2007). "Psychophysiological science: interdisciplinary approaches to classic questions about the mind," in *Handbook of Psychophysiology*, eds J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson (NewYork, NY: Cambridge University Press), 1–16. doi: 10.1017/CBO9780511546396.001

Cassidy, S. M., Robertson, I. H., and O'Connell, R. G. (2012). Retest reliability of event-related potentials: evidence from a variety of paradigms. *Psychophysiology* 49, 659–664. doi: 10.1111/j.1469-8986.2011.01349.x

Causse, M., Chua, Z., Peysakhovich, V., Campo, N., and Matton, N. (2017). Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS. *Sci. Rep.* 7:5222. doi: 10.1038/s41598-017-05378-x

Cavanagh, J. F., and Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends Cogn. Sci.* 18, 414–421. doi: 10.1016/j.tics.2014.04.012

Cegovnik, T., Stojmenova, K., Jakus, G., and Sodnik, J. (2018). An analysis of the suitability of a low-cost eye tracker for assessing the cognitive load of drivers. *Appl. Ergon.* 68, 1–11. doi: 10.1016/j.apergo.2017.10.011

Chen, L. L., Zhao, Y., Ye, P. F., Zhang, J., and Zou, J. Z. (2017). Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers. *Expert Syst. Appl.* 85, 279–291. doi: 10.1016/j.eswa.2017.01.040

Chen, L. L., Zhao, Y., Zhang, J., and Zou, J. Z. (2015). Automatic detection of alertness/drowsiness from physiological signals using wavelet-based nonlinear features and machine learning. *Expert Systems Appl.* 42, 7344–7355. doi: 10.1016/j.eswa.2015.05.028

Cheng, S. Y., Park, S., and Trivedi, M. M. (2007). Multi-spectral and multi-perspective video arrays for driver body tracking and activity analysis. *Comput. Vis. Image Underst.* 106, 245–257. doi: 10.1016/j.cviu.2006.08.010

Cohen, M. X. (2011). It's about time. *Front. Human Neurosci.* 5:2. doi: 10.3389/fnhum.2011.00002

Correa, A. G., Orosco, L., and Laciar, E. (2014). Automatic detection of drowsiness in EEG records based on multimodal analysis. *Med. Eng. Phys.* 36, 244–249. doi: 10.1016/j.medengphy.2013.07.011

Costa, N., Simões, P., Costa, S., and Arezes, P. (2017). "Driving workload indicators: the case of senior drivers," in *International Conference on Applied Human Factors and Ergonomics* (Cham: Springer), 604–615.

Dasari, D., Shou, G., and Ding, L. (2017). ICA-Derived EEG correlates to mental fatigue, effort, and workload in a realistically simulated air traffic control task. *Front. Neurosci.* 11:297. doi: 10.3389/fnins.2017.00297

Dawson, M. E., Schell, A. M., and Filion, D. L. (2007). The electrodermal system. *Handb. Psychophysiol.* 2, 200–223. doi: 10.1017/CBO9780511546396.007

de Waard, D. (1996). *The Measurement of Drivers' Mental Workload*. Netherlands: Groningen University, Traffic Research Center.

de Waard, D., and Lewis-Evans, B. (2014). Self-report scales alone cannot capture mental workload. *Cogn. Technol. Work* 16, 303–305. doi: 10.1007/s10111-014-0277-z

Demberg, V. (2013). "Pupillometry: the index of cognitive activity in a dual-task study," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 35. Available online at: https://escholarship.org/uc/item/4vf5w6bn#author

Drews, F. A., Strayer, D. L., Uchino, B., and Smith, T. W. (2012). "The road to road rage," in *Vision in Vehicles X*, eds A. G. Gale, S. P. Taylor, and C. Castro (Elsevier), 346–352.

Duchowski, A. T., Krejtz, K., Krejtz, I., Biele, C., Niedzielska, A., Kiefer, P., et al. (2018). "The index of pupillary activity: measuring cognitive load vis-à-vis task difficulty with pupil oscillation," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal, QC). doi: 10.1145/3173574.3173856

ElKomy, M., Abdelrahman, Y., Funk, M., Dingler, T., Schmidt, A., and Abdennadher, S. (2017). "ABBAS: an adaptive bio-sensors based assistive system," *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '17* (Denver, CO; New York, NY: ACM), 2543–2550. doi: 10.1145/3027063.3053179

Engström, J., Johansson, E., and Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transp. Res. Part F Traffic Psychol. Behav.* 8, 97–120. doi: 10.1016/j.trf.2005.04.012

Eoh, H. J., Chung, M. K., and Kim, S. H. (2005). Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. *Int. J. Indus. Ergon.* 35, 307–320. doi: 10.1016/j.ergon.2004.09.006

Fabiani, M., Gratton, G., and Federmeier, K. D. (2007). "Event-related brain potentials: methods, theory, and applications," in *Handbook of Psychophysiology*, eds J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson (New York, NY: Cambridge University Press), 85–119. doi: 10.1017/CBO9780511546396.004

Fairclough, S. H., Karran, A. J., and Gilleade, K. (2015). "Classification accuracy from the perspective of the user: real-time interaction with physiological computing," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul), 3029–3038. doi: 10.1145/2702123.2702454

Fallahi, M., Motamedzade, M., Heidarimoghadam, R., Soltanian, A. R., and Miyake, S. (2016). Effects of mental workload on physiological and subjective responses during traffic density monitoring: a field study. *Appl. Ergon.* 52, 95–103. doi: 10.1016/j.apergo.2015.07.009

Faure, V., Lobjois, R., and Benguigui, N. (2016). The effects of driving environment complexity and dual tasking on drivers' mental workload and eye blink behavior. *Transp. Res. Part F Traffic Psychol. Behav.* 40, 78–90. doi: 10.1016/j.trf.2016.04.007

Forsman, P. M., Vila, B. J., Short, R. A., Mott, C. G., and van Dongen, H. P. A. (2013). Efficient driver drowsiness detection at moderate levels of drowsiness. *Accid. Anal. Prev.* 50, 341–350. doi: 10.1016/j.aap.2012.05.005

Fowles, D. C. (1986). The eccrine system and electrodermal activity. *Psychophysiology* 1, 51–96. doi: 10.4236/ojpsych.2013.31007

Fridlund, A. J., and Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology* 23, 567–589. doi: 10.1111/j.1469-8986.1986.tb00676.x

Fu, R., and Wang, H. (2014). Detection of driving fatigue by using noncontact EMG and ECG signals measurement system. *Int. J. Neural Syst.* 24:1450006. doi: 10.1142/S0129065714500063

Fu, R., Wang, H., and Zhao, W. (2016). Dynamic driver fatigue detection using hidden Markov model in real driving condition. *Expert Syst. Appl.* 63, 397–411. doi: 10.1016/j.eswa.2016.06.042

Gade, R., and Moeslund, T. B. (2014). Thermal cameras and applications: a survey. *Mach. Vis. Appl.* 25, 245–262. doi: 10.1007/s00138-013-0570-5

Gagl, B., Hawelka, S., and Hutzler, F. (2011). Systematic influence of gaze position on pupil size measurement: analysis and correction. *Behav. Res. Methods* 43, 1171–1181. doi: 10.3758/s13428-011-0109-5

Gao, Z., Li, S., Cai, Q., Dang, W., Yang, Y., Mu, C., et al. (2018). Relative wavelet entropy complex network for improving EEG-based fatigue driving classification. *IEEE Trans. Instrum. Meas.* 99, 1–7. doi: 10.1109/TIM.2018.2865842

Gateau, T., Durantin, G., Lancelot, F., Scannella, S., and Dehais, F. (2015). Real-time state estimation in a flight simulator using fNIRS. *PLoS ONE* 10:e0121279. doi: 10.1371/journal.pone.0121279

Getzmann, S., Arnau, S., Karthaus, M., Reiser, J. E., and Wascher, E. (2018). Age-related differences in pro-active driving behaviour revealed by EEG measures. *Front. Human Neurosci.* 12:321. doi: 10.3389/fnhum.2018.00321

Gilzenrat, M. S. (2006). *The Role of the Locus Coeruleus in Cognitive Control: Pupillometric Measures and Computational Mechanisms*. 0724–0724. Ph.D. thesis, Princeton University.

Gratton, G., and Fabiani, M. (2001). Shedding light on brain function: the event-related optical signal. *Trends Cogn. Sci.* 5, 357–363. doi: 10.1016/S1364-6613(00)01701-0

Gratton, G., and Fabiani, M. (2003). The event-related optical signal (EROS) in visual cortex: replicability, consistency, localization, and resolution. *Psychophysiology* 40, 561–571. doi: 10.1111/1469-8986.00058

Grinvald, A., Lieke, E., Frostig, R. D., Gilbert, C. D., and Wiesel, T. N. (1986). Functional architecture of cortex revealed by optical imaging of intrinsic signals. *Nature* 324, 361–364. doi: 10.1038/324361a0

Grossman, P. (1992). Respiratory and cardiac rhythms as windows to central and autonomic biobehavioral regulation: selection of window frames, keeping the panes clean and viewing the neural topography. *Biol. Psychol.* 34, 131–161. doi: 10.1016/0301-0511(92)90013-K

Guoping, S., and Zhang, K. (2009). An ERP study of effects of driving fatigue on auditory attention. *Psychol. Sci.* 32, 517–520. Available online at: https://psycnet.apa.org/record/2009-12822-001

Haouij, N. E., Poggi, J.-M., Sevestre-Ghalila, S., Ghozi, R., and Jaïdane, M. (2018). "AffectiveROAD system and database to assess driver's attention," in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing* (Pau). doi: 10.1145/3167132.3167395

Healey, J., Seger, J., and Picard, R. (1999). Quantifying driver stress: developing a system for collecting and processing bio-metric signals in natural situations. *Biomed. Sci. Instrum.* 35, 193–198.

Hebb, D. O. (1955). Drives and the CNS (conceptual nervous system). *Psychol. Rev.* 62:243. doi: 10.1037/h0041823

Heine, T., Lenis, G., Reichensperger, P., Beran, T., Doessel, O., and Deml, B. (2017). Electrocardiographic features for the measurement of drivers' mental workload. *Appl. Ergon.* 61, 31–43. doi: 10.1016/j.apergo.2016.12.015

Herff, C., Putze, F., and Schultz, T. (2017). "Evaluating fNIR-based workload discrimination in a realistic driving scenario," in *The First Biannual Neuroadaptive Technology, Conference* (Berlin).

Hess, E. H., and Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science* 143, 1190–1192. doi: 10.1126/science.143.3611.1190

Hidalgo-Muñoz, A. R., Béquet, A. J., Astier-Juvenon, M., Pépin, G., Fort, A., Jallais, C., et al. (2018). Respiration and heart rate modulation due to competing cognitive tasks while driving. *Front. Human Neurosci.* 12:525. doi: 10.3389/fnhum.2018.00525

Hirao, A., Kato, K., Kitazaki, S., and Yamazaki, N. (2007). *Evaluations of Physical Fatigue During Long-Term Driving with a New Driving Posture*. SAE Technical Paper, SAE World Congress & Exhibition (No. 2007-01-0348). doi: 10.4271/2007-01-0348

Horat, S. K., Herrmann, F. R., Favre, G., Terzis, J., Debatisse, D., Merlo, M. C., et al. (2016). Assessment of mental workload: a new electrophysiological method based on intra-block averaging of ERP amplitudes. *Neuropsychologia* 82, 11–17. doi: 10.1016/j.neuropsychologia.2015.12.013

Hostens, I., and Ramon, H. (2005). Assessment of muscle fatigue in low level monotonous task performance during car driving. *J. Electromyogr. Kinesiol.* 15, 266–274. doi: 10.1016/j.jelekin.2004.08.002

Hung, Y. C., Wang, Y. K., Prasad, M., and Lin, C. T. (2017). "Brain dynamic states analysis based on 3D convolutional neural network," in *Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference* (Banff, AB), 222–227. doi: 10.1109/SMC.2017.8122606

Jacob, S., Bukhari, S. S., Ishimaru, S., and Dengel, A. (2018). "Gaze-based interest detection on newspaper articles," in *Proceedings of the 7th Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction* (Warsaw), 4. doi: 10.1145/3208031.3208034

Jagannath, M., and Balasubramanian, V. (2014). Assessment of early onset of driver fatigue using multimodal fatigue measures in a static simulator. *Appl. Ergon.* 45, 1140–1147. doi: 10.1016/j.apergo.2014.02.001

Janisse, M. P. (1977). *Pupillometry: The Psychology of the Pupillary Response*. Halsted Press.

Jasper, P., Sibley, C., and Coyne, J. (2016). "Using heart rate variability to assess operator mental workload in a command and control simulation of multiple unmanned aerial vehicles," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 60 (Sage CA; Los Angeles, CA: SAGE Publications), 1125–1129. doi: 10.1177/1541931213601264

Jennings, J. R., and Allen, B. (2016). "Methodology," Chapter. in *Handbook of Psychophysiology, 4th Edn.* eds J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, 583–611. *Cambridge Handbooks in Psychology*. Cambridge: Cambridge University Press. doi: 10.1017/9781107415782.027

Jennings, J. R., Bberg, W. K., Hutcheson, J. S., Obrist, P., Porges, S., and Turpin, G. (1981). Publication guidelines for heart rate studies in man. *Psychophysiology* 18, 226–231. doi: 10.1111/j.1469-8986.1981.tb03023.x

Jensen, O., and Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Human Neurosci.* 4:186. doi: 10.3389/fnhum.2010.00186

Jobsis, F. F. (1977). Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science* 198, 1264–1267. doi: 10.1126/science.929199

Johansson, G., Evans, G. W., Cederström, C., Rydstedt, L. W., Fuller-Rowell, T., and Ong, A. D. (2012). The effects of urban bus driving on blood pressure and musculoskeletal problems: a quasi-experimental study. *Psychosom. Med.* 74, 89–92. doi: 10.1097/PSY.0b013e31823ba88f

Jonsson, S., and Jonsson, B. (1975). Function of the muscles of the upper limb in car driving. *Ergonomics* 18, 375–388. doi: 10.1080/00140137508931471

Kahana, M. J. (2006). The cognitive correlates of human brain oscillations. *J. Neurosci.* 26, 1669–1672. doi: 10.1523/JNEUROSCI.3737-05c.2006

Kahneman, D., and Beatty, J. (1966). Pupil diameter and load on memory. *Science* 154, 1583–1585. doi: 10.1126/science.154.3756.1583

Kajiwara, S. (2014). Evaluation of driver's mental workload by facial temperature and electrodermal activity under simulated driving conditions. *Int. J. Automot. Technol.* 15, 65–70.

Kassab, A., Le Lan, J., Tremblay, J., Vannasing, P., Dehbozorgi, M., Pouliot, P., et al. (2018). Multichannel wearable fNIRS-EEG system for long-term clinical monitoring. *Human Brain Mapp.* 39, 7–23. doi: 10.1002/hbm.23849

Kassem, K., Salah, J., Abdrabou, Y., Morsy, M., El-Gendy, R., Abdelrahman, Y., et al. (2017). "DiVA: exploring the usage of pupil diameter to elicit valence and arousal," in *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia* (Stuttgart). doi: 10.1145/3152832.3152836

Käthner, I., Wriessnegger, S. C., Müller-Putz, G. R., Kübler, A., and Halder, S. (2014). Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain–computer interface. *Biol. Psychol.* 102, 118–129. doi: 10.1016/j.biopsycho.2014.07.014

Kato, T., Fujii, T., and Tanimoto, M. (2004). "Detection of driver's posture in the car by using far infrared camera," in *Intelligent Vehicles Symposium, 2004* (Parma). doi: 10.1109/IVS.2004.1336406

Ko, L. W., Komarov, O., Hairston, W. D., Jung, T. P., and Lin, C. T. (2017). Sustained attention in real classroom settings: an eeg study. *Front. Human Neurosci.* 11:388. doi: 10.3389/fnhum.2017.00388

Kolli, A., Fasih, A., Al Machot, F., and Kyamakya, K. (2011). "Non-intrusive car driver's emotion recognition using thermal camera," in *Nonlinear Dynamics and Synchronization (INDS) & 16th Int'l Symposium on Theoretical Electrical Engineering* (Klagenfurt). doi: 10.1109/INDS.2011.6024802

Kosiachenko, E., and Si, D. (2017). Classification of brain activity using brainwaves recorded by consumer-grade EEG devices. *Adv. Mass Data*, 14.

Krol, L. R., Zander, T. O., Jaswa, M., Flascher, O., Snelting, A., and Brouwer, A.-M. (2017). "Online-capable cleaning of highly artefactual EEG data recorded during real driving," in *Proceedings of the 7th Graz Brain-Computer Interface Conference.* doi: 10.3217/978-3-85125-533-1-47

Kurosawa, K., Takezawa, N., Sano, T., Miyamoto, S., Yasushi, M., and Hashimoto, H. (2017). "Drowsiness prediction system for vehicle using capacity coupled electrode type non-invasive ECG measurement," in *System Integration (SII), 2017 IEEE/SICE International Symposium*. doi: 10.1109/SII.2017.8279230

Laborde, S., Mosley, E., and Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research–recommendations for experiment planning, data analysis, and data reporting. *Front. Psychol.* 8:213. doi: 10.3389/fpsyg.2017.00213

Le, A. S., Aoki, H., Murase, F., and Ishida, K. (2018). A novel method for classifying driver cognitive distraction under naturalistic conditions with information from near-infrared spectroscopy. *Front. Human Neurosci.* 12:431. doi: 10.3389/fnhum.2018.00431

Lee, B. G., Chong, T. W., Lee, B. L., Park, H. J., Kim, Y. N., and Kim, B. (2017a). Wearable mobile-based emotional response-monitoring system for drivers. *IEEE Trans. Human Mach. Syst.* 47, 636–649. doi: 10.1109/THMS.2017.2658442

Lee, D. S., Chong, T. W., and Lee, B. G. (2017b). Stress events detection of driver by wearable glove system. *IEEE Sens. J.* 17, 194–204. doi: 10.1109/JSEN.2016.2625323

Lee, H. B., Kim, J. S., Kim, Y. S., Baek, H. J., Ryu, M. S., and Park, K. S. (2007). "The relationship between HRV parameters and stressful driving situation in the real road," in *6th International Special Topic Conference on Information Technology Applications in Biomedicine, 2007* (Tokyo). doi: 10.1109/ITAB.2007.4407380

Lee, K. W., Yoon, H. S., Song, J. M., and Park, K. R. (2018). Convolutional neural network-based classification of driver's emotion during aggressive and smooth driving using multi-modal camera sensors. *Sensors* 18:957. doi: 10.3390/s1804095

Leinonen, V., Kankaanpää, M., Vanharanta, H., Airaksinen, O., and Hänninen, O. (2005). Back and neck extensor loading and back pain provocation in urban bus drivers with and without low back pain. *Pathophysiology* 12, 249–255. doi: 10.1016/j.pathophys.2005.09.004

Lenneman, J. K., and Backs, R. W. (2009). Cardiac autonomic control during simulated driving with a concurrent verbal working memory task. *Human Fact.* 51, 404–418. doi: 10.1177/0018720809337716

Lenneman, J. K., and Backs, R. W. (2010). "Enhancing assessment of in-vehicle technology attention demands with cardiac measures," in *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. doi: 10.1145/1969773.1969777

Li, Z., Chen, L., Peng, J., and Wu, Y. (2017). Automatic detection of driver fatigue using driving operation information for transportation safety. *Sensors* 17:1212. doi: 10.3390/s17061212

Li, Z., Zhang, M., Zhang, X., Dai, S., Yu, X., and Wang, Y. (2009). Assessment of cerebral oxygenation during prolonged simulated driving using near infrared spectroscopy: its implications for fatigue development. *Eur. J. Appl. Physiol.* 107, 281–287. doi: 10.1007/s00421-009-1122-6

Liang, B., and Lin, Y. (2018). Using physiological and behavioral measurements in a picture-based road hazard perception experiment to classify risky and safe drivers. *Transp. Res. Part F Traffic Psychol. Behav.* 58, 93–105. doi: 10.1016/j.trf.2018.05.024

Liang, S. F., Lin, C. T., Wu, R. C., Chen, Y. C., Huang, T. Y., and Jung, T. P. (2006). "Monitoring driver's alertness based on the driving performance estimation and the EEG power spectrum analysis," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference* (Shanghai). doi: 10.1109/IEMBS.2005.1615791

Liang, W. C., Yuan, J., Sun, D. C., and Lin, M. H. (2009). Changes in physiological parameters induced by indoor simulated driving: effect of lower body exercise at mid-term break. *Sensors* 9, 6913–6933. doi: 10.3390/s90906913

Liang, Y., Horrey, W. J., Howard, M. E., Lee, M. L., Anderson, C., Shreeve, M. S., et al. (2017). Prediction of drowsiness events in night shift workers during morning driving. *Accid. Anal. Prev.* doi: 10.1016/j.aap.2017.11.004. [Epub ahead of print].

Lim, W. L., Sourina, O., and Wang, L. P. (2018). STEW: Simultaneous task EEG workload dataset. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 2106–2114. doi: 10.1109/TNSRE.2018.2872924

Lin, C. T., Wu, R. C., Liang, S. F., Chao, W. H., Chen, Y. J., and Jung, T. P. (2005). EEG-based drowsiness estimation for safety driving using independent component analysis. *IEEE Trans. Circuits Syst. I Regular Papers* 52, 2726–2738. doi: 10.1109/TCSI.2005.857555

Littler, W. A., Honour, A. J., and Sleight, P. (1973). Direct arterial pressure and electrocardiogram during motor car driving. *Br. Med. J.* 2, 273–277. doi: 10.1136/bmj.2.5861.273

Liu, T., Pelowski, M., Pang, C., Zhou, Y., and Cai, J. (2016). Near-infrared spectroscopy as a tool for driving research. *Ergonomics 59*, 368–379. doi: 10.1080/00140139.2015.1076057

Liu, T., Saito, H., and Oi, M. (2012). Distinctive activation patterns under intrinsically versus extrinsically driven cognitive loads in prefrontal cortex: a near-infrared spectroscopy study using a driving video game. *Neurosci. Lett.* 506, 220–224. doi: 10.1016/j.neulet.2011.11.009

Liu, Y., and Du, S. (2018). Psychological stress level detection based on electrodermal activity. *Behav. Brain Res.* 341, 50–53. doi: 10.1016/j.bbr.2017.12.021

Liu, Y.-T., Lin, Y.-T., Wu, S.-L., Hsieh, T.-Y., and Lin, C.-T. (2015). "Assessment of mental fatigue: an EEG-based forecasting system for driving safety," in *2015 IEEE International Conference on Systems, Man, and Cybernetics* (Kowloon). doi: 10.1109/SMC.2015.561

Lohani, M., and Isaacowitz, D. M. (2014). Age differences in managing response to sadness elicitors using attentional deployment, positive reappraisal and suppression. *Cogn. Emot. 28,* 678–697. doi: 10.1080/02699931.2013.853648

Lohani, M., Payne, B. R., and Isaacowitz, D. M. (2018). Emotional coherence in early and later adulthood during sadness reactivity and regulation. *Emotion* 18, 789–804. doi: 10.1037/emo0000345

Luck, S. J., and Kappenman, E. S. (2012). *Oxford Handbook of ERP Components*. New York, NY: Oxford University Press.

Luijcks, R., Hermens, H. J., Bodar, L., Vossen, C. J., van Os, J., and Lousberg, R. (2014). Experimentally induced stress validated by EMG activity. *PLoS ONE* 9:e95215. doi: 10.1371/journal.pone.0095215

Lundberg, U., Kadefors, R., Melin, B., Palmerud, G., Hassmén, P., Engström, M., et al. (1994). Psychophysiological stress and EMG activity of the trapezius muscle. *Int. J. Behav. Med.* 354–370. doi: 10.1207/s15327558ijbm0104_5

Luque-Casado, A., Perales, J. C., Cárdenas, D., and Sanabria, D. (2016). Heart rate variability and cognitive processing: the autonomic response to task demands. *Biol. Psychol.* 113, 83–90. doi: 10.1016/j.biopsycho.2015.11.013

Lyman, J. M., McGwin, G. Jr., and Sims, R. V. (2001). Factors related to driving difficulty and habits in older drivers. *Accid. Anal. Prev.* 33, 413–421. doi: 10.1016/S0001-4575(00)00055-5

Ma, Z., Yao, S., and Li, J. (2018). "A study of integrated driver fatigue judging algorithm based on principal component analysis," in *MATEC Web of Conferences, (EDP Sciences)* (Nanjing). doi: 10.1051/matecconf/201817302011

Maccora, J., Manousakis, J. E., and Anderson, C. (2018). Pupillary instability as an accurate, objective marker of alertness failure and performance impairment. *J. Sleep Res.* e12739. doi: 10.1111/jsr.12739

Madan, C. R., Harrison, T., and Mathewson, K. E. (2018). Noncontact measurement of emotional and physiological changes in heart rate from a webcam. *Psychophysiology* 55:e13005. doi: 10.1111/psyp.13005

Marinescu, A. C., Sharples, S., Ritchie, A. C., Sánchez López, T., McDowell, M., and Morvan, H. P. (2018). Physiological parameter response to variation of mental workload. *Hum. Factors* 60, 31–56. doi: 10.1177/0018720817733101

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

Marquart, G., Cabrall, C., and de Winter, J. (2015). Review of eye-related measures of drivers' mental workload. *Procedia Manuf.* 3, 2854–2861. doi: 10.1016/j.promfg.2015.07.783

Marshall, S. P. (2002). "The index of cognitive activity: measuring cognitive workload," in *Human Factors and Power Plants, 2002. Proceedings of the 2002 IEEE 7th Conference* (Scottsdale, AZ). doi: 10.1109/HFPP.2002.1042860

Massar, S. A. A., Wester, A. E., Volkerts, E. R., and Kenemans, J. L. (2010). Manipulation specific effects of mental fatigue: evidence from novelty processing and simulated driving. *Psychophysiology* 47, 1119–1126. doi: 10.1111/j.1469-8986.2010.01028.x

Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., and Ro, T. (2009). To see or not to see: prestimulus α phase predicts visual awareness. *J. Neurosci.* 29, 2725–2732. doi: 10.1523/JNEUROSCI.3963-08.2009

Matthews, G., Wohleber, R., Lin, J., Funke, G., and Neubauer, C. (2019). "Monitoring task fatigue in contemporary and future vehicles: a review," in *Advances in Human Factors in Simulation and Modeling. AHFE 2018. Advances in Intelligent Systems and Computing*, Vol. 780, ed D. Cassenti (Cham: Springer) doi: 10.1007/978-3-319-94223-0_10

Mauss, I. B., and Robinson, M. D. (2009). Measures of emotion: a review. *Cogn. Emot.* 23, 209–237. doi: 10.1080/02699930802204677

Mazzetta, I., Gentile, P., Pessione, M., Suppa, A., Zampogna, A., Bianchini, E., et al. (2018). Stand-alone wearable system for ubiquitous real-time monitoring of muscle activation potentials. *Sensors* 18:6. doi: 10.3390/s18061748

McKendrick, R., Parasuraman, R., Murtza, R., Formwalt, A., Baccus, W., Paczynski, M., et al. (2016). Into the wild: neuroergonomic differentiation of hand-held and augmented reality wearable displays during outdoor navigation with functional near infrared spectroscopy. *Front. Hum. Neurosci.* 10:216. doi: 10.3389/fnhum.2016.00216

Mehler, B., Reimer, B., and Coughlin, J. F. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: an on-road study across three age groups. *Hum. Factors* 54, 396–412. doi: 10.1177/0018720812442086

Mehler, B., Reimer, B., Coughlin, J. F., and Dusek, J. A. (2009). The impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transport. Res. Rec.* 2138, 6–12. doi: 10.3141/2138-02

Mehler, B., Reimer, B., and Wang, Y. (2011). "A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload," in *Proceedings of the 6th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design: Driving Assessment, 2011* (Lake Tahoe, CA), 590–597. doi: 10.17077/drivingassessment.1451

Miller, N. R., Newman, N. J., Biousse, V., and Kerrison, J. B. (2005). *Walsh and Hoyt's Clinical-Neuroophthalmology, 6th Edn*. Philadelphia, PA: Lippincott Williams & Wilkins

Moghaddam, G. K., and Lowe, C. R. (eds.). (2019). "*Ex vivo* biosignatures," in *Health and Wellness Measurement Approaches for Mobile Healthcare*. (Cham: Springer), 51–104. doi: 10.1007/978-3-030-01557-2_3

Morad, Y., Lemberg, H., Yofe, N., and Dagan, Y. (2000). Pupillography as an objective indicator of fatigue. *Curr. Eye Res.* 21, 535–542. doi: 10.1076/0271-3683(200007)2111-ZFT535

Morales, J. M., Diaz-Piedra, C., Rieiro, H., Roca-González, J., Romero, S., Catena, A., et al. (2017). Monitoring driver fatigue using a single-channel electroencephalographic device: a validation study by gaze-based, driving performance, and subjective data. *Accid. Anal. Prev.* 109, 62–69. doi: 10.1016/j.aap.2017.09.025

Morris, D. M., Erno, J. M., and Pilcher, J. J. (2017). "Electrodermal response and automation trust during simulated self-driving car use," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, (Los Angeles, CA: SAGE Publications), 61, 1759–1762. doi: 10.1177/1541931213601921

Mun, S., Whang, M., Park, S., and Park, M. (2017). Effects of mental workload on involuntary attention: a somatosensory ERP study. *Neuropsychologia* 106, 7–20. doi: 10.1016/j.neuropsychologia.2017.08.021

Murai, K., Hayashi, Y., Okazaki, T., Stone, L. C., and Mitomo, N. (2008). "Evaluation of ship navigator's mental workload using nasal temperature and heart rate variability," in *Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference* (Singapore). doi: 10.1109/ICSMC.2008.4811503

Nadeau, C., and Bengio, Y. (2000). "Inference for the generalization error," *in Advances in Neural Information Processing Systems*, 307–313. doi: 10.1023/A:1024068626366

Neshov, N. N., Manolova, A. H., Draganov, I. R., Tonschev, K. T., and Boumbarov, O. L. (2018). Classification of Mental Tasks from EEG Signals Using Spectral Analysis, PCA and SVM. *Cybern. Inf. Technol.* 18, 81–92. doi: 10.2478/cait-2018-0007

Noda, A., Miyaji, M., Wakuda, Y., Hara, Y., and Yasuma, F. (2015). Simultaneous measurement of heart rate variability and blinking duration to predict sleep onset and drowsiness in drivers. *J. Sleep Disord Ther* 4, 2167–0277. doi: 10.4172/2167-0277.1000213

Noordzij, M. L., Dorrestijn, S. M., and van den Berg, I. A. (2017). "An idiographic study into the physiology and self-reported mental workload of learning to drive a car," in *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2016 Annual Conference* (Prague). Available online at: https://research.utwente.nl/en/publications/an-idiographic-study-into-the-physiology-and-selfreported-mental-

Nowara, E., M., Marks, T. K., Mansour, H., and Veeraraghavan, A. (2018). "SparsePPG: towards driver monitoring using camera-based vital signs estimation in near-infrared," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (Salt Lake City, UT). doi: 10.1109/CVPRW.2018.00174

Nunez, P. L., and Srinivasan, R. (2006). *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press. doi: 10.1093/acprof:oso/9780195050387.001.0001

O'Donnell, R. D., and Eggemeier, F. T. (1986). "Workload assessment methodology," in *Handbook of perception and human performance. Vol. 2, Cognitive Processes and Performance*, eds K. R. Boff, L. Kaufman, and J. P. Thomas (New York, NY: Wiley), 1–49.

Or, C. K. L., and Duffy, V. G. (2007). Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement. *Occupational Ergonomics 7,* 83–94. Available online at: https://content.iospress.com/articles/occupational-ergonomics/oer00140

Ottaviani, C., Brosschot, J. F., Lonigro, A., Medea, B., Van Diest, I., and Thayer, J. F. (2016). Hemodynamic profiles of functional and dysfunctional forms of repetitive thinking. *Ann. Behav. Med.* 51, 261–271. doi: 10.1007/s12160-016-9851-3

Palinko, O., Kun, A. L., Shyrokov, A., and Heeman, P. (2010). "Estimating cognitive load using remote eye tracking in a driving simulator," in *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (Austin, TX). doi: 10.1145/1743666.1743701

Paredes, P. E., Ordonez, F., Ju, W., and Landay, J. A. (2018). "Fast & furious: detecting stress with a car steering wheel," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal, QC). doi: 10.1145/3173574.3174239

Patel, M., Lal, S. K., Kavanagh, D., and Rossiter, P. (2011). Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert Syst. Appl. 38,* 7235–7242. doi: 10.1016/j.eswa.2010.12.028

Perrier, J., Jongen, S., Vuurman, E., Bocca, M. L., Ramaekers, J. G., and Vermeeren, A. (2016). Driving performance and EEG fluctuations during on-the-road driving following sleep deprivation. *Biol. Psychol.* 121, 1–11. doi: 10.1016/j.biopsycho.2016.09.010

Pfleging, B., Fekety, D. K., Schmidt, A., and Kun, A. L. (2016). "A model relating pupil diameter to mental workload and lighting conditions," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, CA). doi: 10.1145/2858036.2858117

Piña-Ramírez, O., Valdés-Cristerna, R., Medina-Bañuelos, V., and Yañez-Suárez, O. (2018). "Chapter 7 - P300-based brain-computer interfaces," in *Smart Wheelchairs and Brain-Computer Interfaces*, ed P. Diez (Academic Press), 131–170. Available online at: http://www.sciencedirect.com/science/article/pii/B9780128128923000078

Piotrowski, Z., and Szypulska, M. (2017). Classification of falling asleep states using HRV analysis. *Biocybern. Biomed. Eng.* 37, 290–301. doi: 10.1016/j.bbe.2017.02.003

Piper, S. K., Krueger, A., Koch, S. P., Mehnert, J., Habermehl, C., Steinbrink, J., et al. (2014). A wearable multi-channel fNIRS system for brain imaging in freely moving subjects. *Neuroimage* 85, 64–71. doi: 10.1016/j.neuroimage.2013.06.062

Piquado, T., Isaacowitz, D. M., and Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology* 47, 560–569. doi: 10.1111/j.1469-8986.2009.00947.x

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Popescu, F., Blankertz, B., and Müller, K. R. (2008). Computational challenges for noninvasive brain computer interfaces. *IEEE Intell. Syst.* 23, 78–79. Available online at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.370.8282

Prabhakar, G., Madhu, N., and Biswas, P. (2018). "Comparing pupil dilation, head movement, and eeg for distraction detection of drivers," in *Proceedings of the 32nd International BCS Human Computer Interaction Conference* (Belfast, UK). doi: 10.14236/ewic/HCI2018.69

Putman, P., Verkuil, B., Arias-Garcia, E., Pantazi, I., and van Schie, C. (2014). EEG theta/beta ratio as a potential biomarker for attentional control and resilience against deleterious effects of stress on attention. *Cogn. Affect. Behav. Neurosci.* 14, 782–791. doi: 10.3758/s13415-013-0238-7

Quaresima, V., Bisconti, S., and Ferrari, M. (2012). A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults. *Brain Lang.* 121, 79–89. doi: 10.1016/j.bandl.2011.03.009

Quigley, K. S., and Berntson, G. G. (1996). Autonomic interactions and chronotropic control of the heart: heart period versus heart rate. *Psychophysiology* 33, 605–611. doi: 10.1111/j.1469-8986.1996.tb02438.x

Quintana, D. S., Elstad, M., Kaufmann, T., Brandt, C. L., Haatveit, B., Haram, M., et al. (2016). Resting-state high-frequency heart rate variability is related to respiratory frequency in individuals with severe mental illness but not healthy controls. *Sci. Rep.* 6:37212 doi: 10.1038/srep37212

Rai, M., Maity, T., and Yadav, R. K. (2017). Thermal imaging system and its real time applications: a survey. *J. Eng. Technol.* 6, 290–303. Available online at: http://www.joetsite.com/wp-content/uploads/2017/07/Vol.-62-24-2017.pdf

Rajkowski, J., Kubiak, P., and Aston-Jones, G. (1994). Locus coeruleus activity in monkey: phasic and tonic changes are associated with altered vigilance. *Brain Res. Bull.* 35, 607-616. doi: 10.1016/0361-9230(94)90175-9

Rastgoo, M. N., Nakisa, B., Rakotonirainy, A., Chandran, V., and Tjondronegoro, D. (2018). A critical review of proactive detection of driver stress levels based on multimodal measurements. *ACM Comput. Surv.* 51:88. doi: 10.1145/3186585

Recarte, M. A., and Nunes, L. M. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *J. Exp. Psychol* 6, 31–43. doi: 10.1037/1076-898X.6.1.31

Reilly, J., Kelly, A., Kim, S. H., Jett, S., and Zuckerman, B. (2018). The human task-evoked pupillary response function is linear: implications for baseline response scaling in pupillometry. *Behav. Res. Methods* 1–14. doi: 10.3758/s13428-018-1134-4

Reimer, B., Mehler, B., Coughlin, J. F., Godfrey, K. M., and Chuanzhong, T. (2009). "An on-road assessment of the impact of cognitive workload on physiological arousal in young adult drivers," in *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, (ACM), 115–118. doi: 10.1145/1620509.1620531

Ruscio, D., Bos, A. J., and Ciceri, M. R. (2017). Distraction or cognitive overload? using modulations of the autonomic nervous system to discriminate the possible negative effects of advanced assistance system. *Accid. Anal. Prev.* 103, 105–111. doi: 10.1016/j.aap.2017.03.023

Sahayadhas, A., Sundaraj, K., and Murugappan, M. (2012). Detecting driver drowsiness based on sensors: a review. *Sensors* 12, 16937–16953. doi: 10.3390/s121216937

Saus, E. R., Johnsen, B. H., Eid, J., and Thayer, J. F. (2001). Who benefits from simulator training: Personality and heart rate variability in relation to situation awareness during navigation training? *Comput. Hum. Behav.* 28, 1262–1268. doi: 10.1016/j.chb.2012.02.009

Savchenko, V. V., and Poddubko, S. N. (2018). "Development approach to a method for monitoring of driver's ability of resumption of control over the vehicle by on-board systems in automatic mode," in *IOP Conference Series: Materials Science and Engineering*. doi: 10.1088/1757-899X/386/1/012007

Saxby, D. J., Matthews, G., Hitchcock, E. M., Warm, J. S., Funke, G. J., and Gantzer, T. (2008). "Effect of active and passive fatigue on performance using a driving simulator," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 52, (Los Angeles, CA: Sage Publications), 1751–1755. doi: 10.1177/154193120805202113

Schäfer, A., and Vagedes, J. (2013). How accurate is pulse rate variability as an estimate of heart rate variability? a review on studies comparing photoplethysmographic technology with an electrocardiogram. *Int. J. Cardiol.* 166, 15–29. doi: 10.1016/j.ijcard.2012.03.119

Schier, M. A. (2000). Changes in EEG alpha power during simulated driving: a demonstration. *Int. J. Psychophysiol.* 37, 155–162. doi: 10.1016/S0167-8760(00)00079-9

Schmidt, E., and Bullinger, A. C. (2017). Mitigating passive fatigue during monotonous drives with thermal stimuli: Insights into the effect of different stimulation durations. *Accid. Anal. Prev.* 4575, 30443–30448. doi: 10.1016/j.aap.2017.12.005.

Schmidt, E., Decke, R., Rasshofer, R., and Bullinger, A. C. (2017). Psychophysiological responses to short-term cooling during a simulated monotonous driving task. *Appl. Ergon.* 62, 9–18. doi: 10.1016/j.apergo.2017.01.017

Schmidt, E. A., Schrauf, M., Simon, M., Buchner, A., and Kincses, W. E. (2011). The short-term effect of verbally assessing drivers' state on vigilance indices during monotonous daytime driving. *Transport. Res. Part F* 14, 251–260. doi: 10.1016/j.trf.2011.01.005

Schmidt, E. A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., and Kincses, W. E. (2009). Drivers' misjudgement of vigilance state during prolonged monotonous daytime driving. *Accid. Anal. Prev.* 41, 1087–1093. doi: 10.1016/j.aap.2009.06.007

Schneegass, S., Pfleging, B., Broy, N., Heinrich, F., and Schmidt, A. (2013). "A data set of real world driving to assess driver workload," in *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Eindhoven). doi: 10.1145/2516540.2516561

Schwalm, M., Keinath, A., and Zimmer, H., D. (2008). "Pupillometry as a method for measuring mental workload within a simulated driving task," in *Human Factors for assistance and automation*, eds F. Flemisch, B. Lorenz, H. Oberheid, K. A. Brookhuis, and D. de Waard (Maastricht: Shaker Publishing).

Seo, S. H., Kwak, S. H., Chung, S. C., Kim, H. S., and Min, B. C. (2017). Effect of distraction task on driving performance of experienced taxi drivers. *Asian Biomed.* 11, 315–322. doi: 10.1515/abm-2018-0003

Shaffer, F., and Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Front. Public Health* 5:258. doi: 10.3389/fpubh.2017.00258

Shakouri, M., Ikuma, L. H., Aghazadeh, F., and Nahmens, I. (2018). Analysis of the sensitivity of heart rate variability and subjective workload measures in a driving simulator: the case of highway work zones. *Int. J. Ind. Ergon.* 66, 136–145. doi: 10.1016/j.ergon.2018.02.015

Shapiro, D., Jamner, L. D., Lane, J. D., Light, K. C., Myrtek, M., Sawada, Y., et al. (1996). Blood pressure publication guidelines. *Psychophysiology* 33, 1–12. doi: 10.1111/j.1469-8986.1996.tb02103.x

Sheridan, T., Meyer, J., Roy, S., and Decker, K. (1991). *Physiological and Psychological Evaluations of Driver Fatigue During Long Term Driving*. SAE Technical Paper. doi: 10.4271/910116

Shirazi, A. S., Hassib, M., Henze, N., Schmidt, A., and Kunze, K. (2014). "What's on your mind?: mental task awareness using single electrode brain computer interfaces," in *Proceedings of the 5th Augmented Human International Conference* (Kobe). doi: 10.1145/2582051.2582096

Sibi, S., Ayaz, H., Kuhns, D. P., Sirkin, D. M., and Ju, W. (2016). "Monitoring driver cognitive load using functional near infrared spectroscopy in partially autonomous cars," in *Intelligent Vehicles Symposium (IV), 2016 IEEE* (Gothenburg). doi: 10.1109/IVS.2016.7535420

Simon, M., Schmidt, E. A., Kincses, W. E., Fritzsche, M., Bruns, A., Aufmuth, C., et al. (2011). EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions. *Clin. Neurophysiol.* 122, 1168–1178. doi: 10.1016/j.clinph.2010.10.044

Simonson, E., Baker, C., Burns, N., Keiper, C., Schmitt, O. H., and Stackhouse, S. (1968). Cardiovascular stress (electrocardiographic changes) produced by driving an automobile. *Am. Heart J.* 75, 125–135. doi: 10.1016/0002-8703(68)90123-3

Siren, A., Hakamies-Blomqvist, L., and Lindeman, M. (2004). Driving cessation and health in older women. *J. Appl. Gerontol.* 23, 58–69. doi: 10.1177/0733464804263129

Sirevaag, E. J., Kramer, A. F., Coles, M. G. H., and Donchin, E. (1989). Resource reciprocity: an event-related brain potentials analysis. *Acta Psychol.* 70, 77–97. doi: 10.1016/0001-6918(89)90061-9

Sirois, S., and Brisson, J. (2014). Pupillometry. *Wiley Interdiscip. Rev. Cogn. Sci.* 5, 679–692. doi: 10.1002/wcs.1323

Smallwood, J., and Schooler, J. W. (2006). The restless mind. *Psychol. Bull.* 132, 946–958. doi: 10.1037/0033-2909.132.6.946

Soares, V. C. G., Souza, J. K. S., Ginani, G. E., Pompéia, S., Tierra-Criollo, C. J., and Melges, D. B. (2013). "Identification of drowsiness and alertness conditions by means of spectral F-test applied to pupillometric signals," in *Journal of Physics: Conference Series* (Gothenburg), Vol. 477. doi: 10.1088/1742-6596/477/1/012027

Solís-Marcos, I., and Kircher, K. (2018). Event-related potentials as indices of mental workload while using an in-vehicle information system. *Cogn. Technol. Work* 1–13. doi: 10.1007/s10111-018-0485-z

Solovey, E. T., Zec, M., Garcia Perez, E. A., Reimer, B., and Mehler, B. (2014). "Classifying driver workload using physiological and driving performance data: two field studies," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, ON). doi: 10.1145/2556288.2557068

Spruston, N. (2008). Pyramidal neurons: dendritic structure and synaptic integration. *Nat. Rev. Neurosci.* 9, 206–221. doi: 10.1038/nrn2286

Stemberger, J., Allison, R. S., and Schnell, T. (2010). "Thermal imaging as a way to classify cognitive workload," in *2010 Canadian Conference on Computer and Robot Vision*, *IEEE* (Ottawa, ON). doi: 10.1109/CRV.2010.37

Strayer, D. L., and Drews, F. A. (2007). Multi-tasking in the automobile. *Attention: From theory to practice* 121–133.

Strayer, D. L., Turrill, J., Cooper, J. M., Coleman, J. R., Medeiros-Ward, N., and Biondi, F. (2015). Assessing cognitive distraction in the automobile. *Hum. Factors* 57, 1300–1324. doi: 10.1177/0018720815575149

Stuiver, A., Brookhuis, K. A., de Waard, D., and Mulder, B. (2014). Short-term cardiovascular measures for driver support: increasing sensitivity for detecting changes in mental workload. *Int. J. Psychophysiol.* 92, 35–41. doi: 10.1016/j.ijpsycho.2014.01.010

Stuiver, A., de Waard, D., Brookhuis, K. A., Dijksterhuis, C., Lewis-Evans, B., and Mulder, L. J. M. (2012). Short-term cardiovascular responses to changing task demands. *Int. J. Psychophysiol.* 85, 153–160. doi: 10.1016/j.ijpsycho.2012.06.003

Sugie, R., Arakawa, T., and Kozuka, K (2016). Detection of fatigue in long-distance driving by heart rate variability. *Innovat. Comput. Inform. Control Exp. Lett.* 10, 1553–1559. Available online at: https://www.researchgate.net/profile/Toshiya_Arakawa2/publication/304757070_Detection_of_Fatigue_in_Long-distance_Driving_by_Heart_Rate_Variability/links/5779da3608ae4645d611f3cb.pdf

Task Force of the European Society of Cardiology (1996). Heart rate variability, standards of measurement, physiological interpretation, and clinical use. *Circulation* 93, 1043–1065. doi: 10.1161/01.CIR.93.5.1043

Tozman, T., Magdas, E. S., MacDougall, H. G., and Vollmeyer, R. (2015). Understanding the psychophysiology of flow: a driving simulator experiment to investigate the relationship between flow and heart rate variability. *Comput. Human Behav.* 52, 408–418. doi: 10.1016/j.chb.2015.06.023

Tran, C. C., Yan, S., Habiyaremye, J. L., and Wei, Y. (2017). "Predicting driver's work performance in driving simulator based on physiological indices," in *International Conference on Intelligent Human Computer Interaction*, (Cham: Springer), 150–162.

Trivedi, M. M., Cheng, S. Y., Childers, E. M., and Krotosky, S. J. (2004). Occupant posture analysis with stereo and thermal infrared video: algorithms and experimental evaluation. *IEEE Trans. Vehicul. Technol.* 53, 1698–1712. doi: 10.1109/TVT.2004.835526

Tsukahara, J. S., Harrison, T. L., and Engle, R. W. (2016). The relationship between baseline pupil size and intelligence. *Cogn. Psychol.* 91, 109–123. doi: 10.1016/j.cogpsych.2016.10.001

Tsunashima, H., and Yanagisawa, K. (2009). Measurement of brain function of car driver using functional near-infrared spectroscopy (fNIRS). *Comput. Intell. Neurosci.* 2009:164958. doi: 10.1155/2009/164958

Unni, A., Ihme, K., Jipp, M., and Rieger, J. W. (2017). Assessing the driver's current level of working memory load with high density functional near-infrared spectroscopy: a realistic driving simulator study. *Front. Hum. Neurosci.* 11:167. doi: 10.3389/fnhum.2017.00167

van Boxtel, A. (2001). Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral, and neck muscles. *Psychophysiology* 38, 22–34. doi: 10.1111/1469-8986.3810022

van Gent, P., Farah, H., van Nes, N., and van Arem, B. (2018). Analysing noisy driver physiology real-time using off-the-self sensors: heart rate analysis software from the taking the fast lane project. *J. Open Res. Softw.* doi: 10.13140/RG.2.2.24895.56485

Vicente, J., Laguna, P., Bartra, A., and Bailón, R. (2016). Drowsiness detection using heart rate variability. *Med. Biol. Eng. Comput.* 54, 927–937. doi: 10.1007/s11517-015-1448-7

Villringer, A., Planck, J., Hock, C., Schleinkofer, L., and Dirnagl, U. (1993). Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. *Neurosci. Lett.* 154, 101–104. doi: 10.1016/0304-3940(93)90181-J

Visnovcova, Z., Mestanik, M., Gala, M., Mestanikova, A., and Tonhajzerova, I. (2016). The complexity of electrodermal activity is altered in mental cognitive stressors. *Comput. Biol. Med.* 79, 123–129. doi: 10.1016/j.compbiomed.2016.10.014

Vogelpohl, T., Kühn, M., Hummel, T., and Vollrath, M. (2018). Asleep at the automated wheel—Sleepiness and fatigue during highly automated driving. *Accid. Anal. Prev.* 4575, 30117–30119. doi: 10.1016/j.aap.2018.03.013

von Lühmann, A., Herff, C., Heger, D., and Schultz, T. (2015). Toward a wireless open source instrument: functional near-infrared spectroscopy in mobile neuroergonomics and BCI applications. *Front. Hum. Neurosci.* 9:617. doi: 10.3389/fnhum.2015.00617

Voytek, B., and Knight, R. T. (2015). Dynamic network communication as a unifying neural basis for cognition, development, aging, and disease. *Biol. Psychiatry* 77, 1089–1097. doi: 10.1016/j.biopsych.2015.04.016

Wang, F., Wang, H., and Fu, R. (2018). Real-Time ECG-based detection of fatigue driving using sample entropy. *Entropy* 20:196. doi: 10.3390/e20030196

Wang, H. (2015). "Detection and alleviation of driving fatigue based on EMG and EMS/EEG using wearable sensor," in *Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare*. doi: 10.4108/eai.14-10-2015.2261628

Wang, L., Wang, H., and Jiang, X. (2017). A new method to detect driver fatigue based on EMG and ECG collected by portable non-contact sensors. *PROMET-Traffic Transport.* 29, 479–488. doi: 10.7307/ptt.v29i5.2244

Wang, X., and Chuan, X. (2016). Driver drowsiness detection based on non-intrusive metrics considering individual specifics. *Accid. Anal. Prev.* 95, 350–357. doi: 10.1016/j.aap.2015.09.002

Wanyan, X., Zhuang, D., Lin, Y., Xiao, X., and Song, J. -W. (2018). Influence of mental workload on detecting information varieties revealed by mismatch negativity during flight simulation. *Int. J. Ind. Ergon.* 64, 1–7. doi: 10.1016/j.ergon.2017.08.004

Wascher, E., Arnau, S., Gutberlet, I., Karthaus, M., and Getzmann, S. (2018). Evaluating pro-and re-active driving behavior by means of the EEG. *Front. Hum. Neurosci.* 12:205. doi: 10.3389/fnhum.2018.00205

Wascher, E., Rasch, B., Sänger, J., Hoffmann, S., Schneider, D., Rinkenauer, G., et al. (2014). Frontal theta activity reflects distinct aspects of mental fatigue. *Biol. Psychol.* 96, 57–65. doi: 10.1016/j.biopsycho.2013.11.010

Wei, C. S., Lin, Y. P., Wang, Y. T., Lin, C. T., and Jung, T. P. (2018). A subject-transfer framework for obviating inter-and intra-subject variability in EEG-based drowsiness detection. *Neuroimage* 174, 407–419. doi: 10.1016/j.neuroimage.2018.03.032

Wekselblatt, J. B., and Niell, C. M. (2015). Behavioral state—getting in the zone. *Neuron* 87, 7–9. doi: 10.1016/j.neuron.2015.06.020

Wijsman, J., Grundlehner, B., Penders, J., and Hermens, H. (2013). Trapezius muscle EMG as predictor of mental stress. *ACM Trans. Embedded Comput. Syst.* 12:99. doi: 10.1145/2485984.2485987

Wikström, B.-O. (1993). Effects from twisted postures and whole-body vibration during driving. *Int. J. Ind. Ergon.* 12, 61–75. doi: 10.1016/0169-8141(93)90038-F

Wilhelm, B. J., Widmann, A., Durst, W., Heine, C., and Otto, G. (2009). Objective and quantitative analysis of daytime sleepiness in physicians after night duties. *Int. J. Psychophysiol.* 72, 307–313. doi: 10.1016/j.ijpsycho.2009.01.008

Yakobi, O. (2018). "Determinants of Association and Dissociation between Subjective and Objective Measures of Workload," in *Proceedings of the*

Human Factors and Ergonomics Society Annual Meeting, (Los Angeles, CA: SAGE Publications).

Yan, S., Tran, C. C., Wei, Y., and Habiyaremye, J. L. (2017). Driver's mental workload prediction model based on physiological indices. *Int. J. Occup. Saf. Ergon.* 15, 1–9. doi: 10.1080/10803548.2017.1368951

Yang, S., Hosseiny, S. A. R., Susindar, S., and Ferris, T. K. (2016). "Investigating driver sympathetic arousal under short-term loads and acute stress events," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 60, (Los Angeles, CA: SAGE Publications).

Yarkoni, T., and Westfall, J. (2017). Choosing prediction over explanation in psychology: lessons from machine learning. *Perspect. Psychol. Sci.* 12, 1100–1122. doi: 10.1177/1745691617693393

Yerkes, R. M., and Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *J. Comp. Neurol. Psychol.* 18, 459–482. doi: 10.1002/cne.920180503

Ying, L., Fu, S., Qian, X., and Sun, X. (2011). Effects of mental workload on long-latency auditory-evoked-potential, salivary cortisol, and immunoglobulin A. *Neurosci. Lett.* 491, 31–34. doi: 10.1016/j.neulet.2011.01.002

Yoshida, Y., Ohwada, H., Mizoguchi, F., and Iwasaki, H. (2014). Classifying cognitive load and driving situation with machine learning. *Int. J. Mach. Learn. Comput.* 4:210. doi: 10.7763/IJMLC.2014.V4.414

Yoshino, K., Oka, N., Yamamoto, K., Takahashi, H., and Kato, T. (2013). Functional brain imaging using near-infrared spectroscopy during actual driving on an expressway. *Front. Hum. Neurosci.* 7:882. doi: 10.3389/fnhum.2013.00882

Zander, T. O., Andreessen, L. M., Berg, A., Bleuel, M., Pawlitzki, J., Zawallich, L., et al. (2017). Evaluation of a dry EEG system for application of passive brain-computer interfaces in autonomous driving. *Front. Hum. Neurosci.* 11:78. doi: 10.3389/fnhum.2017.00078

Zhao, C., Zhao, M., Liu, J., and Zheng, C. (2012). Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. *Accid. Anal. Prev.* 45, 83–90. doi: 10.1016/j.aap.2011.11.019

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

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 11 June 2019
doi: 10.3389/fnhum.2019.00191

EEG-Based Prediction of Cognitive Load in Intelligence Tests

Nir Friedman[1,2], *Tomer Fekete*[2], *Kobi Gal*[1,3]* *and Oren Shriki*[2,4,5]*

[1] Department of Software and Information Systems Engineering, Ben-Gurion University of the Negev, Beersheba, Israel, [2] Department of Cognitive and Brain Sciences, Ben-Gurion University of the Negev, Beersheba, Israel, [3] School of Informatics, University of Edinburgh, Edinburgh, United Kingdom, [4] Department of Computer Science, Ben-Gurion University of the Negev, Beersheba, Israel, [5] Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beersheba, Israel

Measuring and assessing the cognitive load associated with different tasks is crucial for many applications, from the design of instructional materials to monitoring the mental well-being of aircraft pilots. The goal of this paper is to utilize EEG to infer the cognitive workload of subjects during intelligence tests. We chose the well established advanced progressive matrices test, an ideal framework because it presents problems at increasing levels of difficulty and has been rigorously validated in past experiments. We train classic machine learning models using basic EEG measures as well as measures of network connectivity and signal complexity. Our findings demonstrate that cognitive load can be well predicted using these features, even for a low number of channels. We show that by creating an individually tuned neural network for each subject, we can improve prediction compared to a generic model and that such models are robust to decreasing the number of available channels as well.

Keywords: brain-computer interface, electroencephalography, cognitive load, machine learning, Raven's matrices

OPEN ACCESS

Edited by:
Tamer Demiralp,
Istanbul University, Turkey

Reviewed by:
Ahmet Omurtag,
Nottingham Trent University, United Kingdom
Pietro Aricò,
Sapienza University of Rome, Italy

***Correspondence:**
Kobi Gal
kobig@bgu.ac.il
Oren Shriki
shrikio@bgu.ac.il

Received: *30 December 2018*
Accepted: *22 May 2019*
Published: *11 June 2019*

Citation:
Friedman N, Fekete T, Gal K and Shriki O (2019) EEG-Based Prediction of Cognitive Load in Intelligence Tests. Front. Hum. Neurosci. 13:191. doi: 10.3389/fnhum.2019.00191

1. INTRODUCTION

The performance of complex tasks requires the integration of various mental resources, such as task-related knowledge, working memory, attention and decision making. However, our brains have limited resources for processing and integrating information. The concept of cognitive load generally refers to the relative load on these limited resources (Sweller et al., 1998; Coyne et al., 2009).

Cognitive workload has been explored from different perspectives. Brouwer et al. (2012) refer to workload as the working memory load in an n-back task. Mills et al. (2017) use simple true-false questions for eliciting low workload and open-ended questions, which require more precise memory, for eliciting high workload. Other studies have emphasized the role of skill acquisition in modeling cognitive load (Sweller et al., 1998). Logan (1985) show that when subjects acquire a skill and learn how to perform a task in an automatic manner, their cognitive workload decreases (Borghini et al., 2017) . Thus, the cognitive load depends not only on task complexity but also on the subject's skill at the given task. A highly complex task performed by a non-skilled individual would result in high cognitive load, whereas a simple task performed by a skilled individual would result in low cognitive load. For example, Stevens et al. (2006) assessed subjects as they were learning to diagnose disorders of organ systems and Mak et al. (2013) focused on performance improvement in a visual-motor task. Both studies showed a decrease in cognitive load metrics, with an increase in task familiarity. In all of these studies, the relative difficulty of the task is seen as a proxy for its associated cognitive load. The difficulty was assessed using a variety of approaches, such as the

type of questions (true-false vs. open ended), subject performance and even participant subjective ratings. A major limitation of many studies is that the levels of difficulty were not rigorously defined. Here, we chose a setting in which problem difficulty was rigorously validated and is commonly used in the psychological literature (see below). Another limitation of previous studies is that cognitive workload was assessed using discrete levels, often only two or three levels (Aricò et al., 2016a,b). In the present study, we use a continuous scale for workload.

In addition to behavioral measures, there is a growing interest in assessing cognitive workload using physiological measures, such as pupil diameter (Palinko et al., 2010). The focus of this paper is on quantifying cognitive workload using measures based on electroencephalography (EEG). Several studies have previously developed EEG-based measures for cognitive load. In particular, it was found that the ratio between the theta power (4–8 Hz) and the alpha power (8–12 Hz), as well as the ratio between the beta power (12–30 Hz) and the alpha power and several related combinations, provided informative indices concerning task engagement and cognitive workload (Pope et al., 1995; Stevens et al., 2006; Mills et al., 2017). Other researchers came to similar conclusions, namely that the relation between different spectral features can help predict cognitive load from EEG (Gerě and Jaušcvec, 1999; McDonald and Soussou, 2011; Conrad and Bliemel, 2016). This study aimed to further expand these studies and develop continuous and more accurate EEG-based measures of cognitive load. Furthermore, we tried to examine the utility of additional measures, in particular network connectivity and signal complexity.

We focused on recording EEG during performance of a well-known psychological assessment tool, the advanced progressive matrices test (Raven, 2000), which is commonly used to measure general intelligence. The test is composed of different problems that involve the manipulation of shapes. Problems are presented to subjects at increasing levels of difficulty. The difficulty of each problem is validated across a large number of subjects (Forbes, 1964; Arthur et al., 1999), in the sense that more difficult questions lead to a higher error rate in the population.

Here, we adopt problem difficulty as the operational definition of cognitive load and demonstrate that it can be predicted from the subject's EEG readings. Specifically, we employ a variety of EEG measures as input to machine-learning algorithms and train them to predict problem difficulty.

As mentioned above, previous studies of EEG-based measures of cognitive load were limited in several ways. In particular, they relied mostly on spectral features and produced a simple discrete measure of either low or high load. In contrast, this paper models cognitive load in a continuous manner. In addition, we go beyond basic spectral features and examine how measures of network connectivity and signal complexity affect the prediction of cognitive load. To measure network connectivity, we used complex network analysis (CNA), which provides measures to examine functional connectivity in the brain (Bullmore and Sporns, 2009; Fekete et al., 2014). Features of neural complexity are often computed using measures of entropy, reflecting the proportion of ordered patterns that can be detected in a signal (Bullmore et al., 2009). To measure neural complexity, we focused on Lempel-Ziv (Tononi and Edelman, 1998) complexity, Multi Scale Entropy (MSE) (Abásolo et al., 2006) and Detrended Fluctuation Analysis (DFA) (Rubin et al., 2013).

The results of this paper demonstrate the applicability of using EEG and machine learning for quantifying cognitive load in well-validated problem-solving tasks. In particular, as EEG and other measures of brain activity become more pervasive, quantitative cognitive load measures could be used to facilitate the design of domains involving real-time problem-solving, such as e-learning, psychometric exams, military training, and more (Ikehara and Crosby, 2005; Mills et al., 2017).

2. METHODS

We recorded EEG from subjects while they solved the Advanced Progressive Matrices set II (Raven test). The 36 problems in the test were presented in increasing levels of difficulty. The raw EEG data were then passed through an artifact removal pipeline (see details below) before extracting EEG-based measures of spectral activity, neural complexity and network connectivity. These measures served as input to machine learning algorithms, which were trained to predict problem difficulty.

2.1. Participants

Fifty-two subjects (26 female and 26 male; age range 21–28, $Mean = 24.55$ years, $SD = 1.76$ years) participated voluntarily in the experiment, provided written informed consent and received compensation for participating. The experiment was approved by the Ben-Gurion University ethics committee. All subjects reported that they are right-handed, have normal or corrected vision, and that they have never completed any sort of intelligence test in the past. Four participants were excluded from the study because they required 10 min or less to solve the entire test or answered correctly 16 problems or less. An additional participant was excluded due to a compromised recording (several electrodes did not record any signal throughout the entire session).

2.2. Experimental Paradigm

Subjects performed the Raven's APM Set II problems (36 items in increasing difficulty level), and instructions were delivered before the test started (see **Figure 1** for an example problem). The test was run with no time limit, with all the key requirements and administration instructions carefully following the manual (Raven et al., 1998). Subjects sat in a comfortable chair facing a computer screen 60 cm away. The test was conducted by displaying the problems on the computer screen (23″, 1,920 × 1,080 resolution, with a 2.3° visual angle between each answer's corners), where the subjects were required to press a keyboard key (with their right hand) in accordance with their chosen answer number. The experiment was programmed in MATLAB (www.mathworks.com, version 2015), using the Psychophysics Toolbox extensions (Brainard and Vision, 1997; Pelli, 1997; Kleiner et al., 2007). Each trial lasted from the presentation of the corresponding problem until subject response, and thus trial duration was variable.

EEG was recorded through the whole session using the g.Tec HIamp system (g.Tec, Austria) with 64 gel-based electrodes

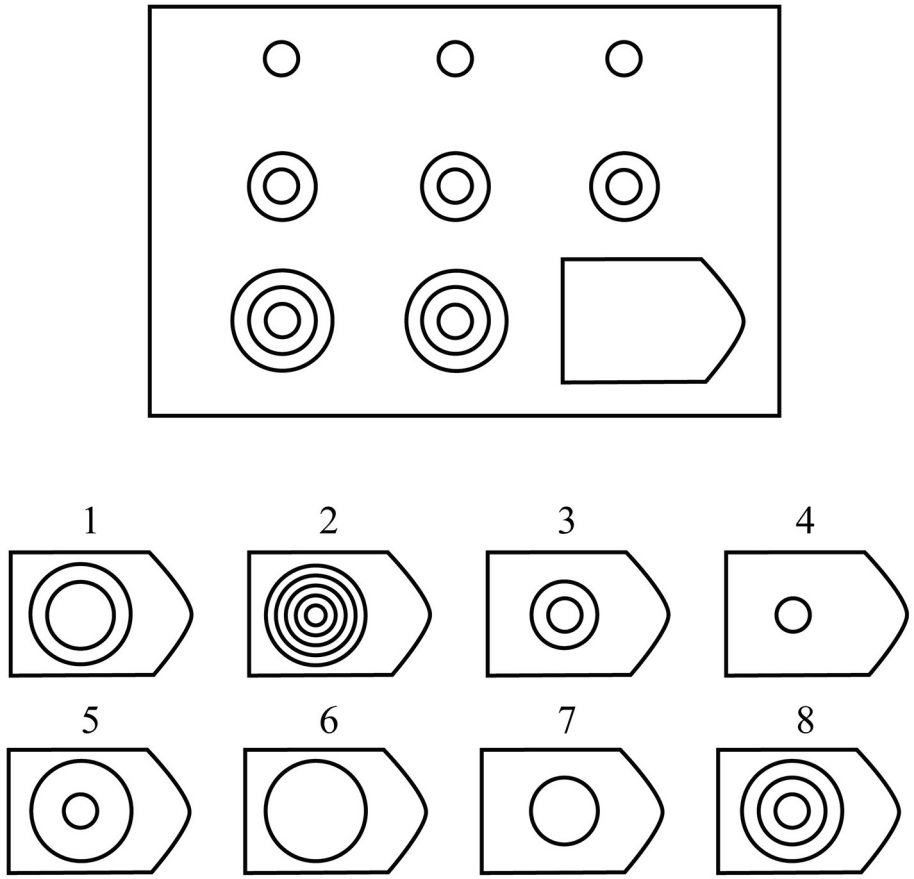


FIGURE 1 | Illustration of Raven's set II Example Problem. The subject is asked to choose the missing shape from the 8 possible options. The correct answer here is option 8.

(AgCl electrolyte gel). Electrodes were positioned according to the standard 10/20 system with linked ears reference. An impedance test and adjustment were carried out at the beginning of the session, and impedances of all electrodes were kept below 5 kΩ. The signal was sampled at 256 Hz with a high-pass filter of 1 Hz. The data were recorded using Matlab Simulink g.Tec plug-ins.

2.3. Feature Extraction

Data were analyzed using a combination of the EEGLAB Matlab toolbox (Delorme and Makeig, 2004) routines and custom code. Data were first high-pass filtered (cut-off 1 Hz), then a customized adaptive filter was applied to suppress line-noise. This was followed by Artifact Subspace Reconstruction (Mullen et al., 2015), re-referencing to the mean, and low-pass filtering (cutoff 60 Hz). Next, Infomax ICA was carried out (Bell and Sejnowski, 1995). The resulting ICs were evaluated automatically for artifacts by combining spatial, spectral and temporal analysis of ICs. ICs identified as containing ocular, muscular or cardiac artifacts were removed from data.

Various features were extracted from the EEG data:

- **Power spectrum metrics (PS)** - The power in 5 frequency bands (delta [1–4 Hz], theta [4–8 Hz], alpha [8–12 Hz], beta [12–30 Hz], and gamma [30–50 Hz]) was calculated for each channel across the whole trial duration. This resulted in 310 features (62 channels × 5 bands) for each trial.
- **Neural complexity metrics**- We focused on three measures of complexity, specifically, Lempel-Ziv complexity (LZC) (Zhang et al., 2001), Multi Scale Entropy (MSE) (Abásolo et al., 2006) and Detrended Fluctuation Analysis (DFA) (Peng et al., 1995; Rubin et al., 2013). The LZC measure was computed as the mean of the measure across all channels, resulting in a single feature for each trial. In comparison, the MSE and DFA measures were first computed for each individual channel, and for DFA we also computed the metric for each frequency band (as described above), and a broadband [1–50Hz]. We then computed the mean, variance, maximum, minimum, $\frac{mean}{variance}$ and $\frac{maximum}{minimum}$, resulting in 6 features for the MSE, and 36 features for DFA [6 measures × (5 bands + 1 broadband)], resulting in 43 complexity features for each trial. Because these metrics are affected by trial duration, we calculated them for the last 2,500 samples (≈10 s) of each trial.
- **Connectivity metrics** - These features are based on a graph reflecting the connectivity of the underlying network. The graph comprises 62 vertices (channels); edges in the graph represent correlations between channels (there are no self

edges). There are two approaches regarding the weight of each edge. One is to take the absolute value of the correlation as the weight of each edge. Another is to give the same weight to all edges that were kept after the thresholding process described below. We kept only the top x% (we tried several thresholds) of the edges with the highest values, for example 5% (which was what we ultimately used), meaning that we were left with 190 edges out of the 62^2 (minus the 62 self edges). The graph was used to extract graph-theoretical features such as average shortest distance between nodes, small-worldness, etc. (Bullmore and Sporns, 2009). We ultimately used the mean and standard deviation of the small-worldness measure and its components, across the different thresholds.

- **Basic** - Simple demographic features of subjects' age and sex were used. In addition, the time it took to answer each problem was used as a feature. These features were added to all the above feature groups in the prediction phase.

3. RESULTS

After removal of subjects who did not meet the inclusion criteria (see Methods), we were left with 47 subjects for the analysis (24 female and 23 male; age range 21–28, *Mean* = 24.55 years, *SD* = 1.79 years). Our goal was to estimate the cognitive workload of subjects as they were trying to solve each problem during the test.

To this end, we assumed that the difficulty level increased with every problem, as validated in previous studies (Forbes, 1964; Arthur et al., 1999). **Figure 2** shows the rate of incorrect responses over all problems in our data, reflecting the established relationship between problem number and difficulty level. Interestingly, problems 24 and 29 deviated significantly from the trend (more than 3 standard deviations). For this reason, both problems were also excluded from our analysis. In addition, we only considered trials where subjects answered correctly. This is based on the premise that the cognitive load exhibited by participants for incorrect answers may not reflect the true level of the question. After excluding the subjects (180 trials) and specific problems (94 trials), as stated above, and the incorrect trials (366 trials), we were left with 1,232 trials.

For each of the 1232 correct trials, we computed different features (as detailed in the Methods) and assigned them with the corresponding difficulty level (a number between 1 and 36) as the target value. Several types of machine learning algorithms were tested in order to predict cognitive load - "Random Forest" (RF) from the sklearn python package (Buitinck et al., 2013) , which is a bagging decision-tree based model (Ho, 1995), and "XGBoost" (XGB) and its corresponding python package (Chen and Guestrin, 2016). XGB is also a decision tree-based model, though it comes from the "boosting" family (Zhou, 2012). They were chosen because of the virtues of an ensemble learning algorithm, along with their usual good fit with temporal data. Additionally, we applied an artificial neural network (ANN), using the keras python package (Chollet et al., 2015). Lastly, we used simple Linear Regression (LR), also from the sklearn python package, as a baseline for comparison. The hyper-parameters of these models were found using a grid search. The best performance was exhibited by the XGBoost classifier with a step size of 0.05. For the optimal feature group, the number of boosting rounds was 300. All other parameters were run with the default settings. All results shown were cross-validated by dividing the data randomly to training and validation sets (80% of the data were used for training, 20% of the data were used for validation) and repeating the process 10–20 times (determined by the time complexity of the analysis). Our main measure of model performance was r^2, which is simply the Pearson correlation squared. It is commonly interpreted as the proportion of the variance for a dependent variable that is explained by an independent variable or variables.

At first, we compared the different feature types in the prediction process with the different classifiers. **Table 1**

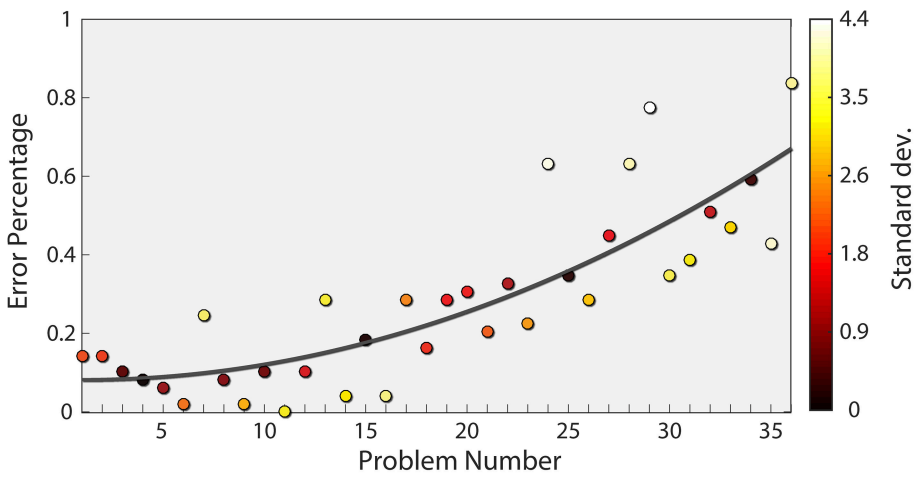


FIGURE 2 | Subject Error rate as a function of problem number. The mean error rate across subjects is plotted for each problem (circles) together with a quadratic fit (dark gray curve). The equation corresponding to the fit is $y = 0.0004793x^2 - 0.000897x - 0.08256$. The color of each point indicates the number of standard deviations from the fit, with bright colors indicating a higher value.

TABLE 1 | The table shows the Pearson correlation (r^2) of each Feature group-Model Pair.

	LR	RF	XGB	ANN
PS	0.007	0.383*	0.655*	0.346*
Complexity	0.323*	0.055	0.508*	0.286*
Connectivity	0.335*	0.186	0.5*	0.267*
PS & Complexity	0	0.322*	0.641*	0.186*
PS & Connectivity	0.07	0.44*	0.67*	0.32*
Complexity & Connectivity	0.339*	0.122	0.519*	0.331*
All Features	0.05	0.358*	0.628*	0.297*

summarizes the r^2 results for this analysis (all results marked with an * were significant). The best results were obtained using XGB for all feature types as seen in a variance test($F(36, 3) = 16.79, p < 0.001$, Tukey multiple comparisons: $p < 0.05$ for all XGB pairs). XGB provides a good trade-off between model complexity and the number of samples required to reach robustness. Even though ANN can capture very complex relationships, they require a large training set. On the other hand, LR and RF do not require significant amounts of training data, but their model complexity is significantly more constrained than XGB.

Next, we compared the utility of each of the three different feature types. PS and connectivity features obtained the highest score, and adding the complexity features to either of the two did not contribute significantly to the prediction. This suggests that complexity features do not add any further information beyond spectral features and connectivity features. To test whether this was not due to high model complexity resulting in over-fitting, we conducted a feature selection process. We found that even after reducing the number of features, no combination of complexity, connectivity and PS features yielded better results than using only the PS and connectivity features together with the basic features.

Figure 3 shows a scatter plot of the best model's prediction together with the true label of each instance in the test set. The Pearson correlation of the best model is $r^2 = 0.67$ ($p < 0.01$). The model was trained on the problem serial number, which should, in principle, produce a linear relationship. However, as evident in **Figure 2**, the relationship between problem number and error rate is slightly non-linear. This suggests that the relationship between problem number and the EEG measure could also be non-linear. We therefore also computed the Spearman correlation, which relates to a general monotonic relationship rather than a linear one, and obtained a value of 0.81 ($p < 0.01$). One of the features used by the algorithm was the duration of each segment, namely the time it took the subject to answer. We also examined the performance with only this feature and found a r^2 of 0.23 ($p < 0.01$) and a Spearman correlation of 0.41 ($p < 0.01$).

3.1. Effect of Number of Electrodes

From an applicative point of view, the number of electrodes affects both the cost and the complexity of using EEG. We therefore examined the extent to which reducing the number of electrodes affects the prediction quality. To this end, we

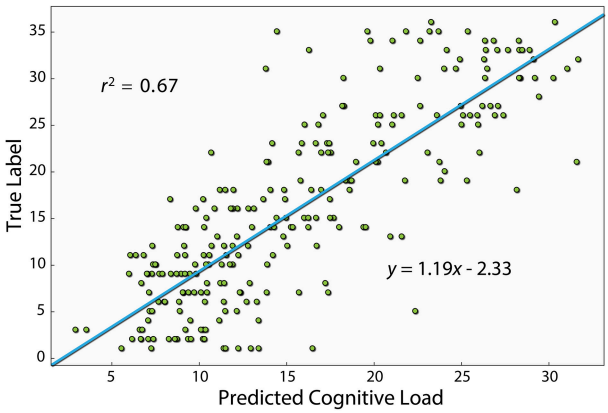


FIGURE 3 | This figure shows the Pearson correlation between the XGBoost model's prediction and the true label of each instance. The model shown here uses the PS, connectivity and basic features, which is the one that produced the best prediction. The equation of the linear fit is $y = 1.19x - 2.33$.

conducted a two step analysis. Firstly, we ran 1,000 simulations, where in each, ten electrodes were chosen randomly out of the total of 62. For each electrode combination, only the relevant PS features were used (five per channel, in addition to the basic features) to generate a workload prediction using the XGB algorithm. We then sorted the electrodes based on the percentage of simulations each electrode was involved in that yielded a score above a specified threshold, out of all simulations it participated in. The top thirty electrodes were chosen in descending order and were taken for the second step, where the effect of the number of best electrodes on the r^2 was examined. As seen in **Figure 4**, a relatively high r^2 of 0.7 ($p < 0.01$) can be obtained using only 12 electrodes (and in fact over 95 percent of peak performance for only 8). Additionally, using the same features of the 12 electrodes, the model produces a Spearman correlation of 0.82 ($p < 0.05$). These 12 electrodes were: CP1, CPz, CP4, TP8, TP10, P3, P4, PO7, O1, O2, AF3, FT8.

3.2. Effect of Discretizing the Workload

In our analysis, the target variable (difficulty of each problem) had 34 possible values. We analyzed the influence of reducing the number of levels of the target variable. We used different sized bins, to reduce the number of different values to 6, 9, 18, 34. For example, to obtain 6 levels, values were binned to [1–6], [7–12], [13–18], [19–24], [25–30], and [31–36]. As evident in **Figure 5**, prediction quality generally decreased with the number of levels. This is not surprising, because the prediction task becomes more complex with the number of levels. In addition, we show that using only the best 12 electrodes found earlier to compute the connectivity features (combined with the PS features of those electrodes), we obtain $r^2 = 0.713$ ($p < 0.05$) for 6 levels, which is the best prediction quality we obtained.

3.3. Individualized Prediction Using Neural Networks

Lastly, because different individuals might experience different levels of cognitive load for the same problem, we wanted to assess

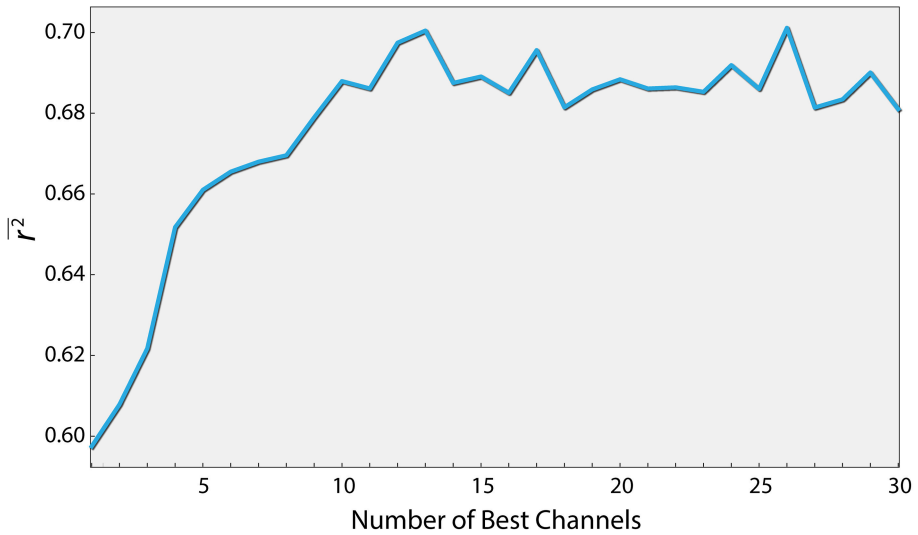


FIGURE 4 | Performance as a function of the number of best channels. Channels were ordered according to their contribution to the prediction quality (see text for details). The curve depicts the prediction quality (r^2) for the XGBoost algorithm as a function of the number of best channels taken into account.

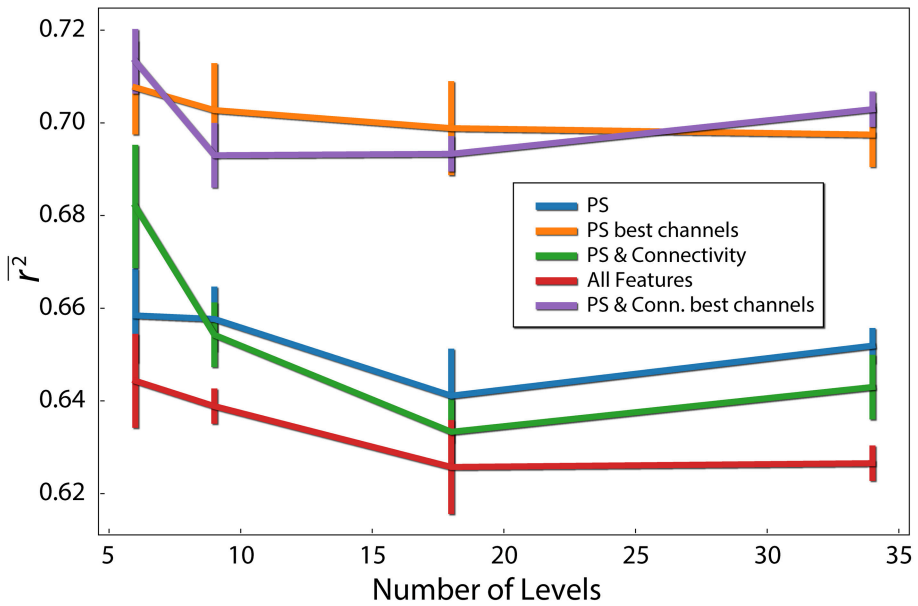


FIGURE 5 | Difficulty level discretization effect on prediction quality (r^2). Each line corresponds to different feature types. PS red are the PS features of the 12 best channels. Error bars reflect standard error of the mean.

the influence of individualizing the prediction model. To this end, we first built a three layer artificial neural network (ANN), trained with data from all subjects using the PS and connectivity features of the 12 best electrodes. We then fixed the parameters of the first and second layers, and for each subject continued to train the weights of the output layer (**Figure 6**). This is a common practice in the field of neural networks (Gruber et al., 2017). We conducted a paired t-test (**Figure 7**), by calculating the mean correlation with the correct answer over several folds using the general model ($M = 0.39, SD = 0.06$) and after tuning ($M = 0.43, SD = 0.06$), which yielded a significant difference ($t = -4.75, p = 0.001$) in favor of the individualized network models.

4. DISCUSSION

We recorded EEG from subjects while solving the advanced progressive matrices test (Raven's matrices test) and used EEG features and machine learning to predict problem difficulty, our chosen operationalization of cognitive workload. Problem

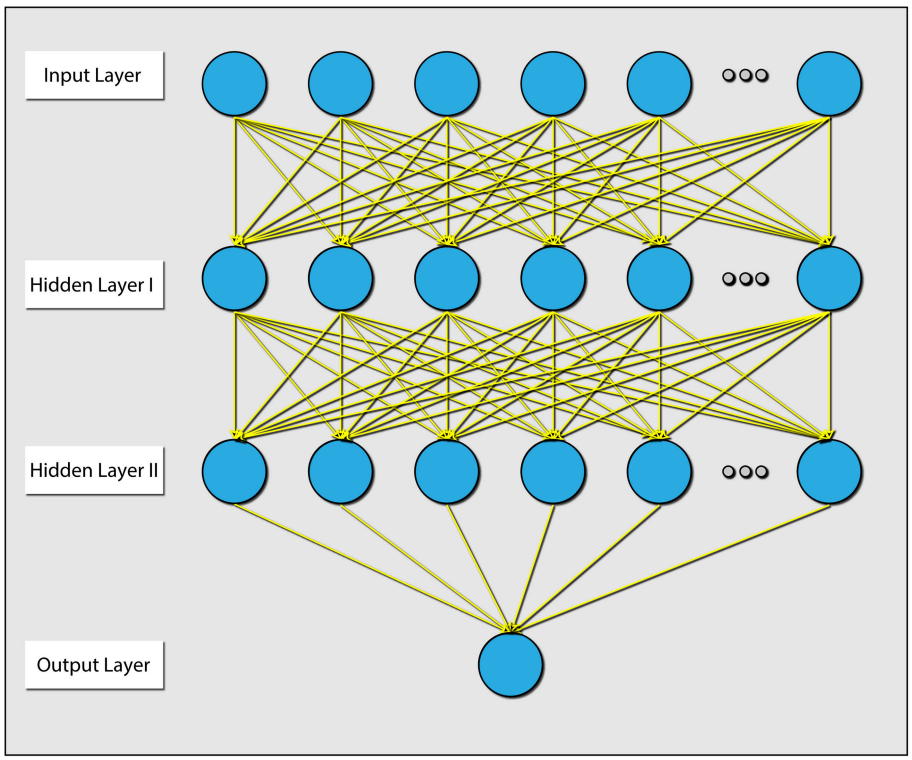


FIGURE 6 | Diagram explaining the architecture of the ANN that was used. There were 2 hidden layers, and all layers were dense (e.g., all connections were present). The parameters between the input layer and hidden layer 1, and the parameters between hidden layer 1 and hidden layer 2 were held during the individualization phase.

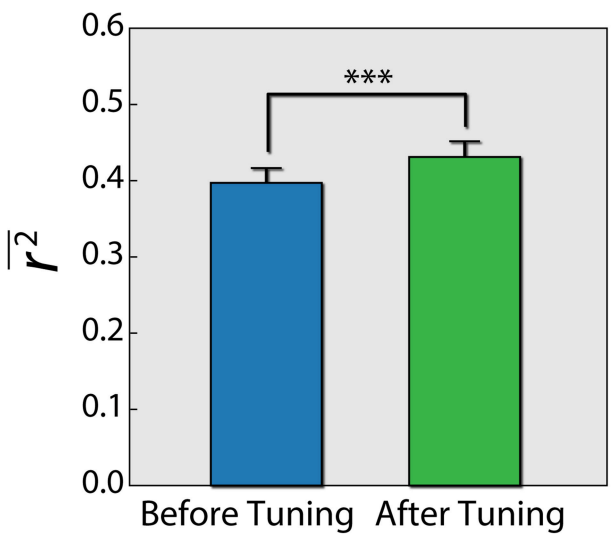


FIGURE 7 | Difference of r^2 score using an ANN before and after individualizing the last layer for each subject. Error bars reflect standard error of the mean. Paired *t*-test results are shown. ***$p = 0.001$.

difficulty was ordered on a scale from 1 to 36 (Forbes, 1964; Arthur et al., 1999) and was treated here as a continuous value. Our results show that even when considering cognitive load in a continuous manner, a reasonable prediction accuracy can be obtained using EEG measures. This could be very useful for many applications in which there is a wide range of cognitive workload levels. These findings extend those of previous studies which used a small number (2-3) of discrete levels of cognitive workload (Gerě and Jaušcvec, 1999; McDonald and Soussou, 2011; Conrad and Bliemel, 2016). Indeed, we found that reducing the number of difficulty levels improves the results significantly.

We examined several machine learning algorithms and found that XGBoost outperformed all other algorithms with all three feature groups. XGBoost was more accurate than the simpler models of linear regression and Random Forest. The lower scores of the ANN are probably due to the fact that they typically require a much larger training dataset than we had at our disposal (Chen and Guestrin, 2016). Furthermore, even though the ANN scored lower than XGBoost, we showed that prediction quality can be improved by tuning the last layer of the ANN to each individual. With a larger dataset, the personalized ANN could potentially attain better prediction than XGBoost. Additionally, in this study we did not use individual features such as individual frequency bands. In general, this could improve the performance of the algorithm. Incorporating individual features should be addressed in future research.

As part of our analysis, we checked the impact of additional EEG measures, specifically metrics of connectivity and metrics of neural complexity. Our results suggest that connectivity measures do add information regarding cognitive load beyond the simple spectral features. On the other hand, it seems

that complexity features, while holding information regarding cognitive load, do not afford additional information over and above that found in connectivity and PS features.

Lastly, we found that prediction quality did not deteriorate, and even improved, when using a limited number of channels ($\sim$ 12), which is important for practical applications. This is most probably due to better generalization, resulting from a less complex model, as opposed to one utilizing all channels.

We chose to utilize the advanced progressive matrices test in this study because of the high validity of its operationalization of difficulty levels. However, to extend our findings further toward applicability, future studies should examine the utility of our EEG-based metrics for cognitive load in real-life settings such as control tower operator performance as aerial traffic ebbs and flows. Since our results indicate the feasibility of employing an array comprising as little as eight electrodes, potentially such studies could be carried out in parallel using portable dry EEG systems. The added benefit would be the feasibility of amassing the expansive datasets necessary for utilizing elaborate neural network models, which in this scenario are expected to improve predictive ability. In addition, it would be useful to identify EEG markers for different dimensions of cognitive workload. Such markers would pave the way for optimizing and personalizing learning processes from e-learning to military training (Ikehara and Crosby, 2005; Mills et al., 2017).

ETHICS STATEMENT

The protocol was approved by the Ben-Gurion University ethics committee. All subjects gave written informed consent in accordance with the Declaration of Helsinki and received compensation for participating.

AUTHOR CONTRIBUTIONS

NF, KG, and OS: experiment planning and design. NF: data acquisition. NF and TF: data analysis. NF, TF, KG, and OS: writing the manuscript.

FUNDING

This research was supported in part by the *Helmsley Charitable Trust* through the Agricultural, Biological and Cognitive (ABC) Robotics Initiative and by the Marcus Endowment Fund both at Ben-Gurion University of the Negev.

ACKNOWLEDGMENTS

The authors thank Dr. Nir Getter for helpful discussions on the manuscript and Ms. Koral Regev for valuable help in running the experiments.

REFERENCES

Abásolo, D., Hornero, R., Gómez, C., García, M., and López, M. (2006). Analysis of eeg background activity in Alzheimer's disease patients with lempel–ziv complexity and central tendency measure. *Med. Eng. Phys.* 28, 315–322. doi: 10.1016/j.medengphy.2005.07.004

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., et al. (2016a). Adaptive automation triggered by eeg-based mental workload index: a passive brain-computer interface application in realistic air traffic control environment. *Front. Hum. Neurosci.* 10:539. doi: 10.3389/fnhum.2016.00539

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Pozzi, S., and Babiloni, F. (2016b). A passive brain–computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks. *Prog. Brain Res.* 228, 295–328. doi: 10.1016/bs.pbr.2016.04.021

Arthur, W. Jr., Tubre, T. C., Paul, D. S., and Sanchez-Ku, M. L. (1999). College-sample psychometric and normative data on a short form of the raven advanced progressive matrices test. *J. Psychoeduc. Assess.* 17, 354–361. doi: 10.1177/073428299901700405

Bell, A. J., and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159. doi: 10.1162/neco.1995.7.6.1129

Borghini, G., Aricò, P., Di Flumeri, G., Cartocci, G., Colosimo, A., Bonelli, S., et al. (2017). Eeg-based cognitive control behaviour assessment: an ecological study with professional air traffic controllers. *Sci. Rep.* 7, 547. doi: 10.1038/s41598-017-00633-7

Brainard, D. H., and Vision, S. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Brouwer, A.-M., Hogervorst, M. A., Van Erp, J. B., Heffelaar, T., Zimmerman, P. H., and Oostenveld, R. (2012). Estimating workload using eeg spectral power and erps in the n-back task. *J. Neural Eng.* 9:045008. doi: 10.1088/1741-2560/9/4/045008

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., et al. (2013). "API design for machine learning software: experiences from the scikit-learn project," in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning* (Prague), 108–122.

Bullmore, E., Barnes, A., Bassett, D. S., Fornito, A., Kitzbichler, M., Meunier, D., et al. (2009). Generic aspects of complexity in brain imaging data and other biological systems. *Neuroimage* 47, 1125–1134. doi: 10.1016/j.neuroimage.2009.05.032

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186. doi: 10.1038/nrn2575

Chen, T., and Guestrin, C. (2016). "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining* (San Francisco, CA: ACM), 785–794. doi: 10.1145/2939672.2939785

Chollet, F. et al. (2015). *Keras*. Available online at: https://keras.io. (accessed May, 2019).

Conrad, C. D., and Bliemel, M. (2016). "Psychophysiological measures of cognitive absorption and cognitive load in E-learning applications," in *Proceedings of the 37th International Conference on Information Systems, December 11–14*, eds P. Agerfalk, N. Levina, and S. S. Kien (Dublin).

Coyne, J. T., Baldwin, C., Cole, A., Sibley, C., and Roberts, D. M. (2009). "Applying real time physiological measures of cognitive load to improve training," in *International Conference on Foundations of Augmented Cognition* (San Diego, CA: Springer), 469–478.

Delorme, A., and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Fekete, T., Beacher, F. D., Cha, J., Rubin, D., and Mujica-Parodi, L. R. (2014). Small-world network properties in prefrontal cortex correlate with predictors of psychopathology risk in young children: a nirs study. *NeuroImage* 85, 345–353. doi: 10.1016/j.neuroimage.2013.07.022

Forbes, A. (1964). An item analysis of the advanced matrices. *Br. J. Educ. Psychol.* 34, 223–236. doi: 10.1111/j.2044-8279.1964.tb00632.x

Gerě, I., and Jaušcvec, N. (1999). Multimedia: Differences in cognitive processes observed with eeg. *Educ. Technol. Res. Dev.* 47, 5–14. doi: 10.1007/BF02299630

Gruber, I., Hlaváč, M., Železnỳ, M., and Karpov, A. (2017). "Facing face recognition with resnet: round one," in *International Conference on Interactive Collaborative Robotics* (Hatfield: Springer), 67–74.

Ho, T. K. (1995). "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, Vol. 1, (Montreal, QC: IEEE), 278–282.

Ikehara, C. S., and Crosby, M. E. (2005). "Assessing cognitive load with physiological sensors," in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 295–303.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., et al. (2007). What's new in psychtoolbox-3. *Perception* 36:1.

Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Can. J. Psychol./Revue Can. Psychol.* 39:367. doi: 10.1037/h0080066

Mak, J. N., Chan, R. H., and Wong, S. W. (2013). "Spectral modulation of frontal eeg activities during motor skill acquisition: task familiarity monitoring using single-channel eeg," in *Engineering in Medicine and Biology Society (embc), 2013 35th Annual International Conference of the IEEE* (Osaka: IEEE), 5638–5641.

McDonald, N. J., and Soussou, W. (2011). "Quasar's qstates cognitive gauge performance in the cognitive state assessment competition 2011," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE* (Boston, MA: IEEE), 6542–6546.

Mills, C., Fridman, I., Soussou, W., Waghray, D., Olney, A. M., and D'Mello, S. K. (2017). "Put your thinking cap on: detecting cognitive load using eeg during learning," in *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (Vancouver, BC: ACM), 80–89.

Mullen, T. R., Kothe, C. A., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., et al. (2015). Real-time neuroimaging and cognitive monitoring using wearable dry eeg. *IEEE Trans. Biomed. Eng.* 62, 2553–2567. doi: 10.1109/TBME.2015.2481482

Palinko, O., Kun, A. L., Shyrokov, A., and Heeman, P. (2010). "Estimating cognitive load using remote eye tracking in a driving simulator," in *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (Austin, TX: ACM), 141–144.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897X00366

Peng, C.-K., Havlin, S., Stanley, H. E., and Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos* 5, 82–87. doi: 10.1063/1.166141

Pope, A. T., Bogart, E. H., and Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biol. Psychol.* 40, 187–195. doi: 10.1016/0301-0511(95)05116-3

Raven, J. (2000). The raven's progressive matrices: change and stability over culture and time. *Cogn. Psychol.* 41, 1–48. doi: 10.1006/cogp.1999.0735

Raven, J. C., and Court, J. H. (1998). *Raven's Progressive Matrices and Vocabulary Scales*. London: Psychological Corporation.

Rubin, D., Fekete, T., and Mujica-Parodi, L. R. (2013). Optimizing complexity measures for fmri data: algorithm, artifact, and sensitivity. *PLoS ONE* 8:e63448. doi: 10.1371/journal.pone.0063448

Stevens, R., Galloway, T., and Berka, C. (2006). "Integrating EEG models of cognitive load with machine learning models of scientific problem solving," in *Augmented Cognition: Past, Present and Future*, eds D. Schmorrow, K. Stanney, and L. Reeves (Arlington, VA: Strategic Analysis), 55–65.

Sweller, J., Van Merrienboer, J. J., and Paas, F. G. (1998). Cognitive architecture and instructional design. *Educ. Psychol. Rev.* 10, 251–296. doi: 10.1023/A:1022193728205

Tononi, G., and Edelman, G. M. (1998). Consciousness and complexity. *science* 282, 1846–1851. doi: 10.1126/science.282.5395.1846

Zhang, X.-S., Roy, R. J., and Jensen, E. W. (2001). Eeg complexity as a measure of depth of anesthesia for patients. *IEEE Trans. Biomed. Eng.* 48, 1424–1433. doi: 10.1109/10.966601

Zhou, Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms*. Raton, FL: Chapman and Hall/CRC. doi: 10.1201/b12207

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

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 09 July 2019
doi: 10.3389/fnhum.2019.00221

A Large-Scale, Cross-Sectional Investigation Into the Efficacy of Brain Training

Adam Hampshire[1]*, *Stefano Sandrone*[1] *and Peter John Hellyer*[2]

[1] *The Computational, Cognitive and Clinical Neuroimaging Laboratory, Division of Brain Sciences, Imperial College London, London, United Kingdom,* [2] *Centre for Neuroimaging Sciences, King's College London, London, United Kingdom*

Brain training is a large and expanding industry, and yet there is a recurrent and ongoing debate concerning its scientific basis or evidence for efficacy. Much of evidence for the efficacy of brain training within this debate is from small-scale studies that do not assess the type of "brain training," the specificity of transfer effects, or the length of training required to achieve a generalized effect. To explore these factors, we analyze cross-sectional data from two large Internet-cohort studies (total $N = 60{,}222$) to determine whether cognition differs at the population level for individuals who report that they brain train on different devices, and across different timeframes, with programs in common use circa 2010–2013. Examining scores for an assessment of working-memory, reasoning and verbal abilities shows no cognitive advantages for individuals who brain train. This contrasts unfavorably with significant advantages for individuals who regularly undertake other cognitive pursuits such as computer, board and card games. However, finer grained analyses reveal a more complex relationship between brain training and cognitive performance. Specifically, individuals who have just begun to brain train start from a low cognitive baseline compared to individuals who have never engaged in brain training, whereas those who have trained for a year or more have higher working-memory and verbal scores compared to those who have just started, thus suggesting an efficacy for brain training over an *extended* period of time. The advantages in global function, working memory, and verbal memory after several months of training are plausible and of clinically relevant scale. However, this relationship is not evident for reasoning performance or self-report measures of everyday function (e.g., employment status and problems with attention). These results accord with the view that although brain training programs can produce benefits, these might extend to tasks that are operationally similar to the training regime. Furthermore, the duration of training regime required for effective enhancement of cognitive performance is longer than that applied in most previous studies.

Keywords: brain training, efficacy of brain training, cross sectional study, memory, commercial brain training

OPEN ACCESS

Edited by:
Michael A. Silver,
University of California, Berkeley,
United States

Reviewed by:
Henry Mahncke,
Posit Science Corporation,
United States
Fred Loya,
Veterans Affairs Northern California
Health Care System (VANCHCS),
United States

***Correspondence:**
Adam Hampshire
a.hampshire@imperial.ac.uk

Received: *05 November 2018*
Accepted: *17 June 2019*
Published: *09 July 2019*

Citation:
Hampshire A, Sandrone S and Hellyer PJ (2019) A Large-Scale, Cross-Sectional Investigation Into the Efficacy of Brain Training. Front. Hum. Neurosci. 13:221. doi: 10.3389/fnhum.2019.00221

INTRODUCTION

Brain training is a large and expanding industry. It has been estimated that sales in this sector are increasing at a compound rate of 20 to 25% annually, passing $1.3bn worldwide in 2013 and projected to exceed $6bn by 2020 (SharpBrains, 2013; Cookson, 2014; Katz, 2014). Brain training has also been the focus of intensive academic research; however, despite the prominence and commercial success of brain training, its efficacy remains the topic of much debate.

Most notably, in 2014 more than 70 scientists signed an open letter entitled "A Consensus on the Brain Training Industry from the Scientific Community," which argued that there is no scientific basis or evidence for the efficacy of brain training (Allaire et al., 2014). In response, 2 months later another group of more than 100 scientists publicly criticized the open letter, in form and substance, claiming that evidence for the "brain training effect" was plentiful and highlighting that the first letter could not be considered a consensus view (Alescio-Lautier et al., 2014). The latter group also accused the former of taking an extreme "faith-based" position, pertaining more to an ideological stance, whilst ignoring the scientific evidence. At present, opinions remain divided.

A number of factors contribute to this controversy. At the most fundamental level there is uncertainty regarding the definition of what exactly constitutes "efficacy". More specifically, the core aim of brain training is to produce general improvements in cognition through repeated exercise on specific computer-based tasks. To be considered effective, brain training should enhance the performance of untrained tasks via improvements in underlying cognitive abilities (Lindenberger et al., 2017). Consequently, validation studies look for evidence of "generalization" or "transfer effects."

Some of the largest academic randomized control trials in computerized cognitive training (ACTIVE, IHAMS and IMPACT) have reported evidence for cognitive improvement, and transfer to everyday cognitive function (e.g., IADLs/HRQoL/depression, IADLs/depression, PROs, respectively) (ACTIVE: Willis et al., 2006; Rebok et al., 2014; IHAMS: Wolinsky et al., 2013; Wolinsky et al., 2016; IMPACT: Smith et al., 2009; Zelinski et al., 2011). However, another large-scale trial published negative results in younger adults (Owen et al., 2010), although positive results, including generalization to real-world measures, were reported for older adults who trained over a longer time frame (Corbett et al., 2015). A recent meta-analyses of cognitive training and a pilot study showed benefits in cognitive function, with the first specifically noting transfer to untrained measures (Mewborn et al., 2017) and the latter reporting short-term functional and long-term structural plastic changes related to gains in global cognition (Lampit et al., 2015a, but see also Lampit et al., 2015b).

However, there is a crucial lack of clarity regarding what "transfer" actually means. In a prominent review, a differentiation between "near" and "far" transfer has been advocated (Simons et al., 2016). Specifically, "near transfer" refers to improvements that generalize to tasks that are operationally similar to the training paradigm; for example, training on one spatial working memory task and observing improvements on another spatial working memory task. In contrast, "far transfer" refers to improvements that generalize more broadly, for example, training on a spatial working memory task and observing improvements in selective attention or a composite construct such as IQ.

Indeed, to "match the hype" of the brain training sector, transfer should not only be "far", but also ecologically valid, namely evident as improvements in everyday function. Seeking to achieve this is quite ambitious. As noted by Simons, there is "no evidence for broad-based improvement in cognition, academic achievement, professional performance, and social competencies that derive from the decontextualized practice of cognitive skills devoid of domain-specific content" (Simons et al., 2016). These broad abilities may rely on factors that brain training regime often neglect, including complex environments offering practice and engagement with domain-related challenges (Simonton, 1990; Shimamura et al., 1995; Staudinger and Baltes, 1996; Stern, 2002; Ericsson, 2006; Rohwedder and Willis, 2010; Grossmann et al., 2012). It is perhaps not surprising that only rare examples of studies reporting "far" transfer effects exist, and most of these studies used children as participants (Thorell et al., 2009; Steiner et al., 2011; Johnstone et al., 2012; Foy and Mann, 2014; Graziano and Hart, 2016; Conklin et al., 2017).

Conversely, evidence for "near transfer" is more convincing. A brain training regime was reported to improve processing speed and executive functions in the elderly (Nouchi et al., 2012) and in young adults (Nouchi et al., 2013). Substantial effects have been reported within the working memory domain for tasks that are similar to the training paradigms (Melby-Lervåg and Hulme, 2013; Karbach and Verhaeghen, 2014; Au et al., 2016; Melby-Lervåg and Hulme, 2016; Soveri et al., 2017; Strobach and Huestegge, 2017). For example, it has been shown that transfer may occur when the category of stimulus is changed and the operational requirements of the paradigm remain similar, but not when the paradigm is changed (Holmes et al., 2018). This lack of far transfer in the context of significant near transfer has also been demonstrated in a population with mild cognitive impairment (Vermeij et al., 2016). Nonetheless, some brain training studies have even failed to find even "near transfer" effects (Guye and von Bastian, 2017).

One might argue that this lack of reproducibility relates to the prevalence of too many parallel trials conducted at small cohort scale. Thus, the academic field of "brain training" has a high risk of type 1 and 2 errors. Notable exceptions to this rule are studies that have measured transfer effects in thousands of individuals. However, even there, the reported results appear contradictory, with some articles claiming significant transfer effects at large scale (Hardy et al., 2015) whereas others have reported negligible transfer even to operationally similar tasks (Owen et al., 2010).

Notably, Owen et al. (2010) have been criticized for providing insufficient "intensity" during the training phase. This criticism warrants further discussion because it highlights an often-overlooked problem: it remains unclear what the optimal parameters for a brain training regime are. Should the training last minutes or hours per session? How many times per week? What timescale should the training program be run

for to produce transfer effects (near or far) of significant scale? Should brain training be paired with physical activity and social interaction to increase the positive effect of the brain training (Boot and Kramer, 2014). This issue relates to a lack of exploratory "scoping" work in the field; evidence from controlled trials forms the ultimate target of intervention research, yet this is often undertaken without prior exploration of study design parameters, which in turn inflates the risk of insensitive and underpowered studies. Brain training-wise, a gap between existing theories and existing data has very recently been highlighted (Edwards et al., 2018). While the option of dismissing effective behavioral interventions on theoretical grounds is not beneficial to public health (Edwards et al., 2018), further investigations are needed before wide implementation of brain training programs. Indeed, it is notable that older adults from the cohort of Owen et al. (2010) did show transfer effects, but they also trained over a longer period of time.

Here we attempt to bridge this knowledge gap with an exploratory cross-sectional investigation of data from two large-scale Internet-cohort studies. In the first cohort, the questionnaire included the question "do you brain train." In the second cohort, we expanded significantly on this question in order to probe intensity, device and length of training, whilst also exploring how these factors might compare with other cognitive pursuits such as gaming. Out hypothesis was that brain training has significantly scaled transfer effects over long but not short time scales. To seek evidence of near transfer, we test whether individuals who used brain training programs in common use in 2010–2013 had a significant advantage in their working-memory, reasoning and verbal scores. We examine how these differences in scores interact with how long participants had been brain-training, i.e., for individuals who had just started to brain train compared to those who do not brain train at all and those who had trained for weeks, months or years. We then assess how cognitive performance varies as a function of training frequency. We also search for evidence of far transfer by comparing employment status and self-reported problems of attention in everyday life across the brain training groups. Finally, we test the hypothesis, that for both near and far transfer, engaging in brain training is as or more effective than alternative cognitive pursuits, including card games, video games and puzzles.

MATERIALS AND METHODS

Cognitive Tasks

The cognitive tasks reported in this study were programmed in Adobe Flex 3 by AH. They have been reported in several previous studies (such as Owen et al., 2010; Hampshire et al., 2012; Daws and Hampshire, 2017) and were adapted for the Internet from classical paradigms in the experimental psychology and cognitive neuroscience literature. They measure planning, reasoning, attention, and working memory abilities. Tasks were presented on a bespoke web-site in a fixed sequence, after which we performed a detailed, demographic assessment. An entire battery of tasks took approximately 30 min to complete, with each task calculating one outcome measure.

Participants

Data collection for Cohort 1 was performed between September and December 2010 via a website advertised in a New Scientist feature, on the Discovery Channel website, in the Daily Telegraph, and on social networking websites including Facebook and Twitter (for further details, please refer to Hampshire et al., 2012). Cohort 2 used a slightly different subset of tasks and was collected in the first four months of 2013 with advertisement through a press release associated with the article published from analysis of Cohort 1 (Hampshire et al., 2012).

In Cohort 1, we included participants who completed all 12 tasks in the analysis (44,780 participants, **Table 1**). In Cohort 2, we included in the analysis only participants who had completed 12 or more of the 13 tasks (15,442).

Ethical approval for the study protocol was awarded by the Cambridge Psychology Research Ethics Committees (2010.62) and the University of Western Ontario Health Sciences Research Ethics Board (103472) for Cohorts 1 and 2, respectively. All participants gave informed consent by clicking a button on the website before being able to access the cognitive and demographic assessment.

Data Analysis

MATLAB and SPSS were used to conduct statistical analyses. The studies were not pre-registered, and the analyses are exploratory rather than resulting from an *a priori* analyses plan. Data from both studies were preprocessed using the following steps:

(i) Participants with ages below 15 or above 90 and subjects with nonsensical responses to any survey question were excluded case-wise (see Hampshire et al., 2012 for further details). Each participant was issued with a username and login. They were able to undertake the tasks multiple times if they wished; however, only their first attempt at the testing battery was analyzed in this study. Individuals who answered the questionnaire too quickly to have read the questions were excluded.

(ii) The cognitive data for each task were ranked and transformed to normality, an approach that deals with Non-normally distributed data and outlier values.

TABLE 1 | Demographics for Cohort 1 (N = 44,780).

Age range (years)	Mean	30
	SD	11,48
Gender	Female	11,633
	Male	33,147
Handedness	Left	5,411
	Right	39,369
Brain train?	Yes	2,833
	No	41,947
Video games?	Daily	12,415
	Weekly	11,911
	Monthly	9,452
	Never	11,002

(iii) Latent variables were estimated separately from the Cohort 1 and Cohort 2 performance data in a data-driven manner using principal component analysis (PCA) as follows.

To define a "Global" measure of cognitive performance in each cohort, we first performed a principal component analysis, on the rank-transformed scores for each of the 12 tasks. The first unrotated principal component was used to define a "global" measure. Mathematically, this is the biggest linear mixture of all abilities that tasks involve and is analogous to an IQ score. To enable finer-grained analysis across different cognitive domains, we defined three orthogonal "summary" variables using a varimax rotation of the PCA coefficients. These latent variables are fully characterized in previous work (Hampshire et al., 2012; Daws and Hampshire, 2017). In brief, significant components were defined using the Kaiser convention, which only includes components that have eigenvalues that are higher than 1 (**Table 2**). In both datasets, three "significant" components were identified. Inspection of the task-component loadings after varimax rotation showed that these summary variables correspond to the working memory (WM), reasoning and verbal demands of the tasks. Multiple abilities underlie performance of each task, and this has been reported extensively in our previous papers. For the sake of consistency with previous studies, we used PCA with varimax rotation and not alternative methods such as PFA. We noted though, that the latter generates a near identical task-factor loading matrix.

(iv) Latent variable scores were generated for the participants using regression. Relationships between latent variable scores and questionnaire variables were determined by generalized linear modeling after factoring out other potentially confounding questionnaire variables.

Analyzing data with large numbers of samples affords very high statistical power, which means that effects of potentially negligible or small scale can have very low p values; therefore, in big-data studies of this type, a more informative gauge of significance is effect size. Here, we conform to Cohen's notion of effect sizes, whereby an effect of ∼0.2 standard deviations (SDs) is small, ∼0.5 SDs is medium, and ∼0.8 SDs is large. All statistical values from our analyses are $p < 0.001$ unless otherwise indicated. All results and figures are presented in SD units, enabling visual assessment of effect size.

TABLE 2 | Brain training and computer gaming vs. task scores in Cohort 1.

		Wald Chi-square	*df*	Sig.
Global score	Video games	1413.65	3	< 0.001
	Brain training	9.98	1	0.002
WM	Video games	608.00	3	< 0.001
	Brain training	14.25	1	< 0.001
Reasoning	Video games	909.80	3	< 0.001
	Brain training	4.18	1	0.041
Verbal	Video games	18.63	3	< 0.001
	Brain training	3.10	1	0.079

RESULTS

Cohort 1 – Is Brain Training Effective?

Brain Training Is Effective, but the Effect Is Small to Negligible When Compared to Regular Video-Gaming

Of the 44,780 participants included in Cohort 1, 2,833 reported that they regularly used a brain training program (**Table 1**). The global measure explained ∼28% of the population variance in performance. The three varimax rotated principal components (**Figure 1**), collectively accounting for ∼46% of the variance. Potentially confounding variables including age, gender, handedness, ethnicity, education level and employment status were factored out of these summary variables prior to further analysis. A general linear model was run including the factors Brain Training (answer "yes" vs. "no" to the question "Do you brain train?") and Video Games (answer "Never," "Monthly," "Weekly" or "Daily" to question "How often do you play Video Games?") with global performance as the predicted variable. The Wald Chi Squared showed statistically significant main effects of Brain Training ($X = 9.98$ $p = 0.002$) and of Video Games ($X = 1413.7$ $p < 0.001$). However, the Brain Training main effect was of small scale (+0.06 SDs). The Video Games main effect was of medium scale and there was a clear relationship with frequency of gaming, specifically, Non-gamers scored 0.47 SDs lower than those who reported playing Video Games daily. Repeating these analyses for the WM, Reasoning and Verbal summary variables (**Table 2** and **Figure 2**) showed negligible scaled main effects for Brain Training. There were significantly scaled main effects for Video Games for the WM and Reasoning variables, but not for the Verbal variable (0.31, 0.37 and 0.024 SDs, respectively). In a final analysis, the scale of the Brain Training effect was examined separately for each age decade. None of the age groups showed a significantly scaled effect, with the largest being for people in their 30 and 60 s (both ∼0.15 SDs).

Cohort 2 – What Are the Factors Affecting Brain Training and Other Cognitive Pursuits?

A plausible explanation for the lack of relationship between cognitive performance and brain training in Cohort 1 was that lower than average cognitive ability motivates people to engage in brain training. If this was the case, then individuals who have just started to brain train would have lower than average task performance, which would mask any benefits. A related possibility was that training may be required at high frequency produce a generalized effect. Furthermore, some training software packages may be more beneficial than others. To explore these possibilities, Cohort 2 completed a more detailed questionnaire, which included the questions "Do you believe that brain training works?," "How often do you brain train?," "How long have you been brain training?" and "Which brain training devices do you use?." There also were questions pertaining to the frequency of other common cognitive pursuits including Video Games, card games, board games, and puzzles such as Sudoku and crossword puzzles.

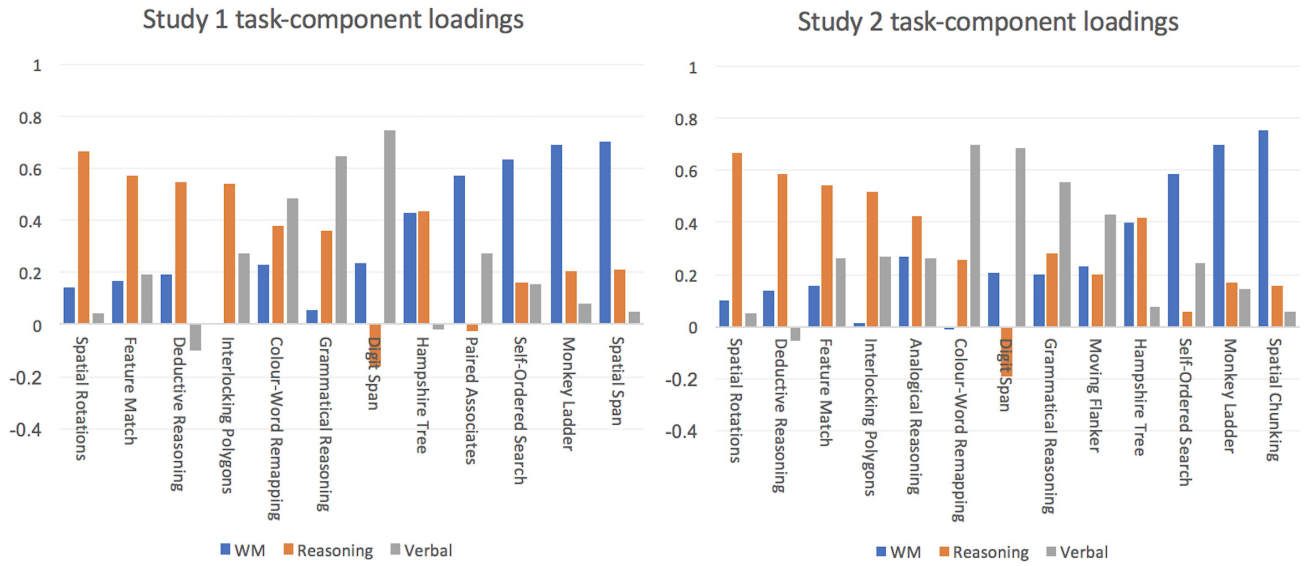


FIGURE 1 | Principle components analysis. Similar varimax rotated 3-component models were evident in Study 1 and 2. One component (WM) explained substantial variance in tasks that require information to be maintained actively in working memory. Another component (Reasoning) explained variance in tasks that required either information to be transformed according to rules (e.g., Rotations and Spatial Planning) or rules to be identified (e.g., Deductive Reasoning). The final component (Verbal) explained variance in tasks that have language or number stimuli.

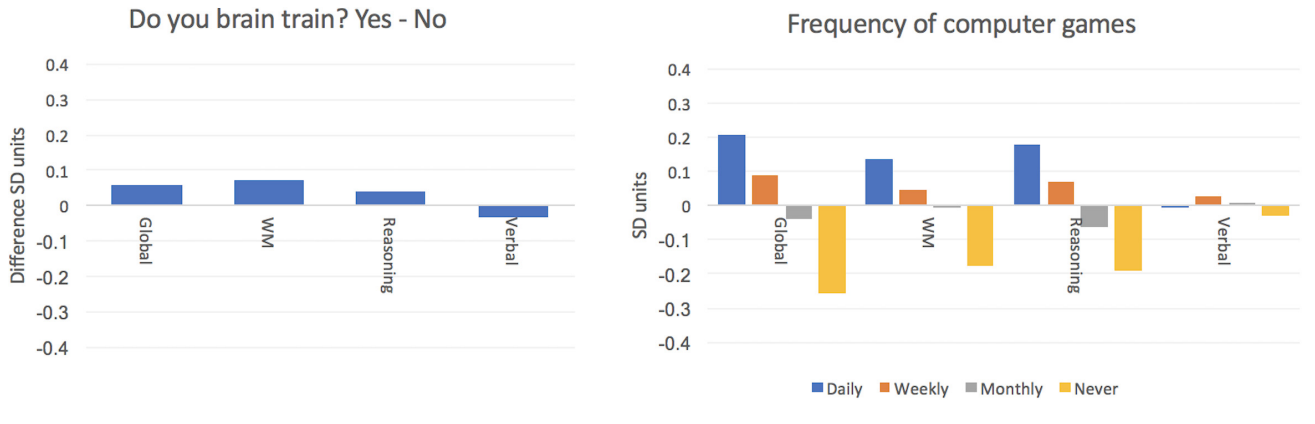


FIGURE 2 | Relationship of brain training and computer gaming with cognitive score in Study 1. Left – In Study 1 there was little difference in cognitive scores for individuals who report regular brain training vs. the rest of cohort. Right – Participants who played Video Games showed small-medium scaled advantages in cognitive scores. These scaled with frequency of gaming and were evident for the Global, WM and Reasoning scores, but not for Verbal score.

Belief in Brain Training Is Consistent With Generalized Strength of Belief

After data cleaning, 15,442 individuals were included in Cohort 2, 3,917 of whom reported that they brain trained (**Table 3**). Approximately half (8,387) of the cohort answered "yes" to the question "Do you believe that brain training works," 1,368 answered "no" with the remaining 5,682 reporting that they did not have an opinion. Interestingly, strength of belief in brain training scaled linearly with strength of religious belief (**Figure 3**). The global performance variable accounted for 27% of the variance in performance. The three varimax rotated components collectively accounted for 43% of the variance (**Figure 1**). Potentially confounding effects of age, handedness, gender, ethnicity, education level, employment status and religious group were factored out of the summary variables prior to further analysis.

Brain Training May Be Effective, but So Are Other Cognitive Pursuits

A general linear model was run with global performance as the predicted variable and including factors for frequency (Never, Monthly, Weekly, Daily) of pursuits including Brain Training, Video Games, Board games, Cards and Puzzles (e.g., crosswords and Sudoku) (frequencies in **Table 4**). All factors showed statistically significant main effects at $p < 0.001$ (**Table 5** and **Figure 4**). The largest positive effect sizes were for Video Games

TABLE 3 | Demographics for Cohort 2 (N = 15,442).

Age range (years)	Mean	26
	SD	12.7
Gender	Female	4,756
	Male	10,683
Handedness	Left	1,638
	Right	13,804
Brain training works?	Yes	8,387
	Maybe	5,682
	No	1,368

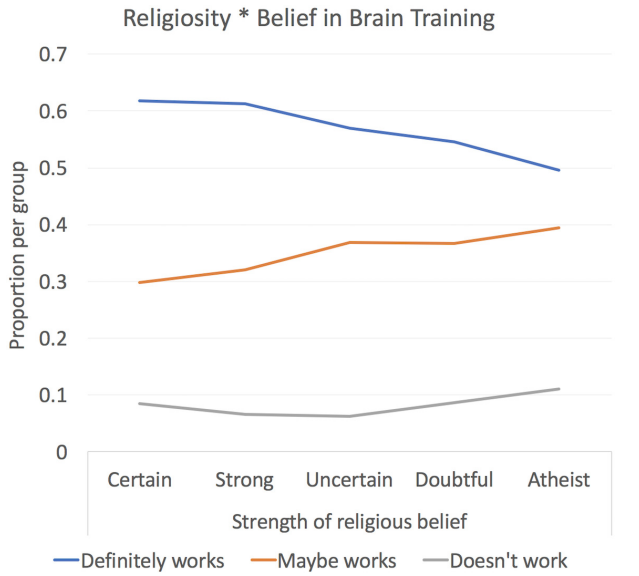


FIGURE 3 | Relationship between religiosity and belief in brain training in Study 2. The majority of participants were of the opinion that brain training "works." However, those who held strong religious beliefs also were more likely to believe in brain training.

TABLE 4 | Frequency of cognitive pursuits.

	Daily	Weekly	Monthly	Never
Brain training	810	1,055	2,052	11,519
Video games	3,572	3,666	3,684	4,515
Card games	441	1,312	6,218	7,466
Board games	265	1,071	6,110	7,991
Puzzles	1,304	4,968	6,428	2,276

at 0.27 SDs and Puzzles at 0.39 SDs. Individuals who reported brain training daily showed a small disadvantage relative to those who did not (e.g., Daily training vs. Never = –0.21 SDs).

At a finer grain, Video Game and puzzle players showed small advantages for the Reasoning variable (0.29 and 0.24 SDs, respectively), Cards and puzzle players showed small advantages for the WM variable (both 0.21 SDs), and puzzle players showed a small advantage for the Verbal variable (0.21 SDs) whereas individuals who brain trained showed a small disadvantage for the Verbal variable (–0.25 SDs).

TABLE 5 | Cognitive pursuits vs. task scores in Cohort 2.

		Wald Chi-square	*df*	Sig.
Global score	Brain training	41.19	3	< 0.001
	Video games	177.29	3	< 0.001
	Card games	18.47	3	< 0.001
	Board games	51.43	3	< 0.001
	Puzzles	358.30	4	< 0.001
WM	Brain training	1.85	3	0.604
	Video games	32.78	3	< 0.001
	Card games	57.20	3	< 0.001
	Board games	24.52	3	< 0.001
	Puzzles	130.83	4	< 0.001
Reasoning	Brain training	17.78	3	< 0.001
	Video games	182.73	3	< 0.001
	Card games	4.25	3	0.235
	Board games	12.65	3	0.005
	Puzzles	127.00	4	< 0.001
Verbal	Brain training	65.48	3	< 0.001
	Video games	10.52	3	0.015
	Card games	4.72	3	0.194
	Board games	26.00	3	< 0.001
	Puzzles	68.20	4	< 0.001

There Are Small Differences Between Common Devices and Packages for Brain Training

When global variable scores were compared for the most common training software packages in the cohort, these being Lumosity (N = 877) Nintendo Brain Age (N = 298) vs. all others. There was no significant main effect of device (p = 0.537). Repeating this analysis at a finer grain showed no statistically significant main effect of device for the WM variable (p = 0.165). There were statistically significant main effects of device for the Reasoning and Verbal variables (p = 0.007 and p = 0.001, respectively). However, these were of negligible effect size, with Brain Age scoring 0.15 SDs higher than Lumosity for the Reasoning variable and Lumosity scoring 0.18 SDs higher than brain Age for the verbal variable.

Frequency and Intensity Are Independent Factors That Contribute to the Efficacy of Brain Training

Individuals who brain trained were examined at an even finer grain by dividing the population into groups according to whether they reported training for a year or more (875), months (704), weeks (695), or had just started (1644). A general linear model was run with global performance as the predicted variable and the factors Training Frequency (Daily, Weekly Monthly) and Training Duration, and the 2-way interaction of these factors (see **Table 6** for cross tabulation). Both main effects were significant at $p < 0.001$. The interaction was statistically Non-significant, which is notable given the statistical power afforded at this cohort scale.

New "Brain Trainers" Start at a Lower Baseline in Cognitive Performance

Examining the data for those individuals who had just started brain training showed that they were on average numerically

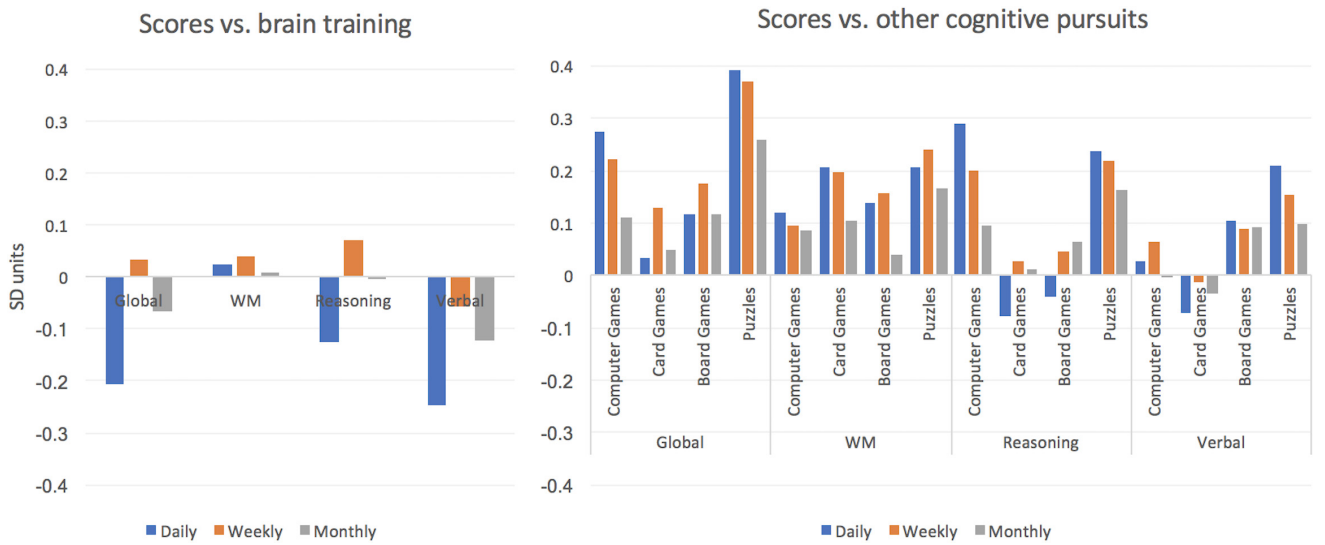


FIGURE 4 | Relationship of brain training and other cognitive pursuits with cognitive scores in Study 2. Left – Cognitive scores for participants who brain train at different frequencies in cohort 2. All measures are relative to those who do not brain train. Participants who engaged in daily brain training showed a small but significant disadvantage in their Global and Verbal scores. Right – Scores broken down according to other cognitive pursuits. All values relative to participants who do not participate in the relevant cognitive pursuit. Small-medium scaled advantages in cognitive scores were evident. These often scaled with frequency. The relationships also varied according to the type of cognitive pursuit. E.g., participants who played card games regularly showed advantages for WM score only whereas those who played Video Games showed advantages for WM and Reasoning but not Verbal scores. Puzzles were associated with higher scores for all three cognitive variables.

below the mean performance of the broader cohort (**Tables 6**, **7** and **Figure 5**). This effect was most pronounced for those individuals who trained on a daily basis (–0.24 SDs). Global performance tracked upward in a linear manner for all three frequency groups as a function of training duration with the highest performing group being those who trained on a weekly basis for > 1 year. This group performed 0.32 SDs higher than the population average for individuals who do not brain train. Repeating this analysis for each composite performance variable showed significant main effects of frequency and duration for the Verbal variable, a significant main effect of duration for the WM variable and a significant main effect of frequency for the Reasoning variable. There were no other significant main effects or interactions (**Table 7** and **Figure 5**).

Brain training has a negligible effect on self-report of everyday problems

Finally, we examined whether the relationship that was evident between brain training duration and cognitive performance extended to every-day life, "far transfer." First, the frequency of self-reported problems concentrating in everyday life was examined (never, infrequently, weekly, several times a week, every day, all the time) for individuals who had brain trained for different time spans. There was a statistically significant main effect of timespan ($P < 0.001$); however, although the group with lowest self-reported scores for problems concentrating were those who had brain trained the longest, the difference relative to those who just started was of negligible scale (0.072 SDs) as was the difference relative to those who do not brain train (0.13 SDs). Then, the proportion of individuals who were employed was examined as a function of time spent brain training. Calculating the strength of association between time spent training (never, just started, weekly, monthly, > 1 year) and employment status

TABLE 6 | Cross tabs for training frequency and training duration.

	Just started	Weeks	Months	>1 year	Total
Daily	240	130	143	298	811
Monthly	1064	316	322	350	2052
Weekly	340	249	239	227	1055
Total	1644	695	704	875	3918

TABLE 7 | Main effects and interactions of frequency and duration.

		Wald Chi-square	*df*	Sig.
Global score	Frequency	28.833	2	< 0.001
	Duration	35.414	3	< 0.001
	Interaction	2.343	6	0.886
WM	Frequency	0.678	2	0.713
	Duration	12.393	3	0.006
	Interaction	5.875	6	0.437
Reasoning	Frequency	19.56	2	< 0.001
	Duration	4.461	3	0.216
	Interaction	4.232	6	0.645
Verbal	Frequency	16.739	2	< 0.001
	Duration	25.729	3	< 0.001
	Interaction	11.723	6	0.068

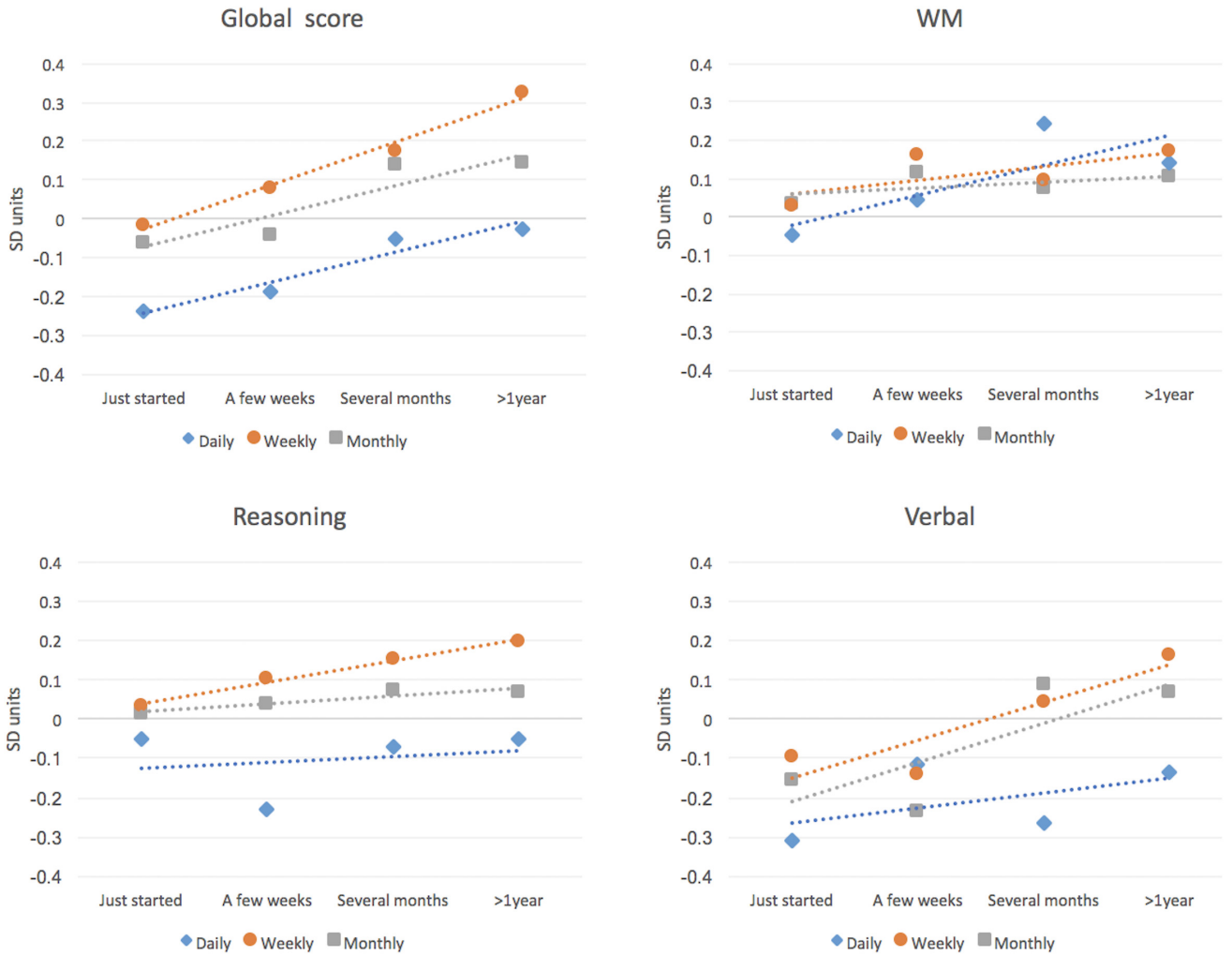


FIGURE 5 | Cognitive scores in Study 2 for people who report brain training at different frequencies and over different durations. Participants who had just started brain training showed significantly scaled disadvantages in Global and Verbal score relative to participants who reported no brain training. These lower scores were most pronounced for participants who reported brain training on a daily basis. There was an increase in cognitive scores with duration of training such that those who trained weekly for a year or more had Global scores 0.32 SDs higher than the non-training population. Smaller scaled trends in the same direction were evident for the WM and Reasoning variables.

(full time, part time and unemployed) again showed a statistically significant but negligible-scaled association (Cramer's V = 0.05, $p < 0.001$) (**Figure 6**).

DISCUSSION

Our large-scale cross-sectional analyses provide population-level insights into the likely efficacy of different brain training programs when applied at different intensities and temporal scales. The findings can help not only in the evaluation of previously reported results, but also in the design of future trials (Seitz, 2017).

At a first pass, we found little cross-sectional evidence for beneficial effects of brain training. More specifically, analysis of performance scores in Cohort 1 showed no advantage for people who brain train vs. those who do not in terms of global or any of the three summary variables. The same was essentially the case for Cohort 2; however, the extended questionnaire allowed this null finding to be examined in more detail.

The most notable finding from this finer grained analysis was that people who have just started to brain training (i.e., such that there was no time to have gained any benefits) tend to have a disadvantage relative to the broader population. Scores then track upward as a function of how long but not how frequently they have trained for. These low scores are best accounted for by motivational factors. Simply put, lower cognitive ability is likely a motivating factor for engaging in brain training. In accordance with this view, far more people believe that training works than engage in it. Furthermore, individuals who train the most frequently show the lowest starting baseline. Accounting for

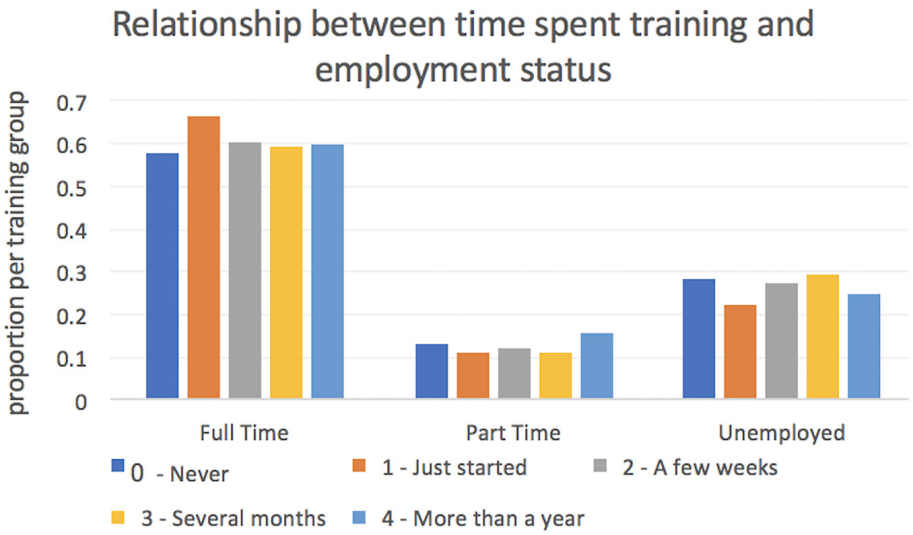


FIGURE 6 | Relationship between time spent brain training and employment in Study 2. Approximately 70% of the study cohort reported being in full time employment. There was no significantly scaled relationship between employment status and the reported duration of brain training.

the differential baseline and its relationship to population variability in motivational factors is an important consideration for any future studies.

The higher performance observed in those who train for longer durations is more promising. It could well be the case that although those who engage in brain training start from a lower than average cognitive baseline, through practice they are able to improve. It is important to note that our results are from cross sectional data and must be corroborated through longitudinal trials. Nonetheless, a synergy exists between observational research and controlled trials, with the former helping to provide a guide to the focus and design of the latter, whilst the latter provides evidence for cause-effect relationships that cannot be directly inferred from the former. In the case of brain training, we argue that there has been insufficient observational research, leading to suboptimal design in many published brain training studies. Few studies have sought to determine the training timespan that is required to produce transfer effects. The guidance of observational studies can help inform optimal parameter ranges. The large-scale observational study presented here provides novel insights that may guide the design of future trials. Most notably, the significantly scaled differences in cognitive performance within the brain training group were observed when comparing those who had been training for a year or more with those who had just started. On the one side, many among the previous studies have operated at substantially shorter timescales: findings from training studies conducted at the scale of weeks or months should be treated cautiously, especially when conducted in small cohorts. On the other side, the differences between the two groups could be explained by selective attrition (e.g., people with low cognitive function are not able to sustain brain training for several months or a year), although our study is not suitable to directly assess this aspect.

Our analyses also showed no interaction or main effect of brain training frequency. On the surface, this provides little support for high intensity regimes, which, from a pragmatic perspective, is important if longer time scales are required. Although previously published reports have suggested a relationship between training frequency and the scale of transfer effects (cited in the review by Simons et al., 2016), the extent and the significance of this relationship remains elusive. *De facto*, this relationship is often assessed retrospectively, which is an aspect directly linked to the nature of the experimental study design used, and the impact of frequency on the size of the transfer may hide the contribution of other factors. A possible explanation for this can thus be the greater sensitivity and control afforded by longitudinal within-subject designs. Moreover, another explanation encompasses the fact that there may be an interaction between baseline ability and rate of improvement with training (higher ability individuals may tend to learn faster) or even an important role of other factors such as motivation. These factors may effectively cancel each other out, thus nullifying their effects. Future controlled intervention trials will assess whether intensity of training plays a key role and will clarify the relationship between training frequency and the scale of the transfer effect. Regardless, such trials should be conducted at longer duration. In line with this interpretation, the study by Corbett et al. (2015) showed cognitive and functional improvements in older adults with a brain training program over a longer timescale (Corbett et al., 2015). A possible complication with longer regime training is that dropout rates may be higher for individuals who start from lower cognitive baselines. This could conceivably produce the illusion of an improvement for groups who had been training for longer time spans. However, we consider this to be unlikely, especially taking into account that previous longitudinal studies have observed no systematic bias in compliance for high vs.

low performing individuals (Hardy et al., 2015); if anything, it is the lower performing individuals who are the most motivated to engage in the training in this study as well, at least as it can be gauged by frequency, although it could still be the case that one's perceived improvement is itself a motivating factor.

A key issue pertains to how far the benefits of brain training generalize. In their piece, Simons and colleagues suggested that benefits likely transfer to cognitive tasks that are similar to the training paradigms. However, evidence for "far" transfer to operationally distinct tasks or improvement in everyday cognitive function after brain training is lacking (Simons et al., 2016; Lindenberger et al., 2017). Our results accord well with this view. For example, there were subtle differences in the relationship between scores for the WM, Reasoning and verbal variables and training software package, which likely relates to the different composition of paradigms that are used.

More importantly, we found little evidence of far transfer effects at the population level. Specifically, when we examined two ecologically focused self-report measures: frequency of problems concentrating in everyday life and employment status. These also showed no overall relationship with brain training and correlation with training frequencies and timescale were of small scale. Again, this result accords well with Simons' perspective on the limited scope of transfer (Simons et al., 2016). It also highlights the importance of assessing the ecological relevance of transfer effects when designing brain training studies.

More tentatively, it could be the case that to achieve generalization to everyday function in clinical populations, it will be necessary to develop training regimes that are closely targeted to the specific operational impairments that contribute to the problems they have in everyday tasks. It may also be advisable to couch such training in a more ecologically relevant format, i.e., by designing training with real-world applications (Moreau and Conway, 2014), virtual environments or augmented reality with similarities to the everyday tasks that the individual would most benefit from improving at. Also, as everyday life involves interaction with other people, ecological validity should take into account factors linked to social interactions (Engert et al., 2017; Valk et al., 2017).

The comparison of these results for brain training to other pursuits is informative in terms of simple mean differences to controls, namely those who do not brain train. Solving puzzles such as crosswords or jigsaws is of course cognitively challenging (Fissler et al., 2017) as are a vast range of other cognitive pursuits. The relationship between frequency of video games, board games and puzzles and cognitive performance were all of significant scale relative to Non-engagement. It is interesting to note that the relationships were not homogenous, i.e., different cognitive pursuits correlated with advantages in different cognitive domains. this again accords with the notion that if there are generalized benefits of engaging in such pursuits, then they likely extend in "near" as opposed to "far" manner.

We cannot rule out the possibility that these relationships have a basis in motivational factors: motivation is very different for cognitive pursuits such as computer gaming, because these are undertaken for entertainment, those who engage in such pursuits may be more motivated to do so if they are more cognitively able and perform better. However, such robust relationships warrant further attention in future studies, with a cross sectional focus extending to baseline performance differences, and further empirical work focusing on carefully controlled "game-training" trials.

Indeed, the current literature on "video-game training" is analogous to that for brain training. For example, some studies have reported significant generalized benefits (Basak et al., 2008; Boot et al., 2008; Anguera et al., 2013; Granic et al., 2014; Mayas et al., 2014; Toril et al., 2014; Green and Bavelier, 2015; Bediou et al., 2018), whereas others have no or only modest benefits (Unsworth et al., 2015; Ballesteros et al., 2017; Sala et al., 2018), with some tentative meta-analytic evidence for both near and far transfer (Wang et al., 2016; Bediou et al., 2018). Once again, the timescales required for generalized benefits is poorly defined and may underlie this inconsistency. The need for larger cohort studies and more intervention studies with more than 30 h of training has already been argued (Bediou et al., 2018), as well as the importance of considering the role of motivational effects in order to rule out alternative explanations before attributing the effect to interventions (Foroughi et al., 2016).

A final interesting point pertains to the relationship between religiosity and belief in the efficacy of brain training. It is intriguing that, as discussed above, those researchers who hold favorable opinions of brain training accused those who do not of taking a faith based position in their open letter. We have previously published analysis of religiosity and its relationship to other variables. It is somewhat ironic that it is, in fact, religious individuals, who are characterized by faith based decisions (Daws and Hampshire, 2017), those who are most likely to believe that "brain training works," and it might be that some people are likely to believe in claims presented to them quite generally. Moreover, belief in different contexts correlates, although there might be other potential confounds (e.g., geography, age, SES). This has implications for where purveyor of brain training technology may best target their products.

Our study has many strengths, but it is important to mention the limitations that might affect the interpretation of our main findings, such as the retrospective nature of report, its Non-experimental design and potential biases inherent to self-report. Our finding that brain training has a negligible effect without long term practice is complicated by the fact that in this cross-sectional analysis participants who underwent brain-training had a heterogenous experience (i.e., focused on a range of domains). This adds noise to our findings, and potentially deflates the scale of our inference and may play a role in explaining the null results here presented, compounded by the retrospective style of the analysis, and the self-report of brain training by participants. Assessing the relationships between the content of brain training and the specific outcomes can be considered. Naturally assessing relationships between the content of brain training, and specific cognitive outcomes should be considered in future works. However, it must be noted that analysis in an

interventional experimental design would require more focused parameters for duration and frequency as identified here. On a more general level, the differences between the brain training groups who have just started brain training and those that have trained for more than a year could be explained by selective attrition (e.g., people with low cognitive function are not able to sustain brain training for several months or a year).

In conclusion, we provide cross-sectional results from two large Internet-based cohorts that accord with the view that individuals who undertake commercially available brain training regimes for long timescales gain benefits that transfer in a limited way to other computerized tasks. Motivation to engage in brain training is shown to be an important confounding factor because it correlates with baseline cognitive ability. Other types of cognitive pursuit are associated with greater performance advantages in the general population and warrant further investigation with controlled trials. Future trials aimed at validating training regimes should focus on longer time-spans, carefully control for baseline ability and motivational factors, and quantify transfer to everyday function. Clinical applications of training should focus on cognitive operations that form the specific basis of patients' impairments in order to minimize transfer distance to everyday function.

ETHICS STATEMENT

Ethical approval for the study protocol was awarded by the Cambridge Psychology Research Ethics Committees (2010.62) and the University of Western Ontario Health Sciences Research Ethics Board (103472) for Cohorts 1 and 2, respectively. All participants gave informed consent by clicking a button on the website before being able to access the cognitive and demographic assessment.

AUTHOR CONTRIBUTIONS

AH and PH made substantial contributions to the conception and design of the work, as well as to the acquisition, analysis and interpretation of data for the work, drafted the work and critically revised it for important intellectual content. SS made substantial contributions to the interpretation of data for the work and critically revised the work for important intellectual content. All authors provided approval for publication of the content and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work were appropriately investigated and resolved.

REFERENCES

Alescio-Lautier, B., Allen, M., Andersen, R. A., Ball, K. K., Banai, K., Baniel, A., et al. (2014). *Cognitive Training Data Response Letter*. Available at: https://www.cognitivetrainingdata.org/the-controversy-does-brain-training-work/response-letter/ (accessed June 19, 2018).

Allaire, J. C., Bäckman, L., Balota, D. A., Bavelier, D., Bjork, R. A., Bower, G. H., et al. (2014). *A Consensus on the Brain Training Industry from the Scientific Community*. Available at: http://longevity.stanford.edu/a-consensus-on-the-brain-training-industry-from-the-scientific-community-2/ (accessed June 19, 2018).

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., et al. (2013). Video game training enhances cognitive control in older adults. *Nature* 501, 97–101. doi: 10.1038/nature12486

Au, J., Buschkuehl, M., Duncan, G. J., and Jaeggi, S. M. (2016). There is no convincing evidence that working memory training is NOT effective: a reply to Melby-Lervåg and Hulme (2015). *Psychon. Bull. Rev.* 23, 331–337. doi: 10.3758/s13423-015-0967-4

Ballesteros, S., Mayas, J., Prieto, A., Ruiz-Marquez, E., Toril, P., and Reales, J. M. (2017). Effects of video game training on measures of selective attention and working memory in older adults: results from a randomized controlled trial. *Front. Aging Neurosci.* 9:354. doi: 10.3389/fnagi.2017.00354

Basak, C., Boot, W. R., Voss, M. W., and Kramer, A. F. (2008). Can training in a real-time strategy video game attenuate cognitive decline in older adults? *Psychol. Aging* 23, 765–777. doi: 10.1037/a0013494

Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., and Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychol. Bull.* 144, 77–110. doi: 10.1037/bul0000130

Boot, W. R., and Kramer, A. F. (2014). The Brain-Games conundrum: does cognitive training really sharpen the mind? *Cerebrum* 2014:15.

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., and Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychol.* 129, 387–398. doi: 10.1016/j.actpsy.2008.09.005

Conklin, H. M., Ashford, J. M., Clark, K. N., Martin-Elbahesh, K., Hardy, K. K., Merchant, T. E., et al. (2017). Long-term efficacy of computerized cognitive training among survivors of childhood cancer: a single-blind randomized controlled trial. *J. Pediatr. Psychol.* 42, 220–231. doi: 10.1093/jpepsy/jsw057

Cookson, C. (2014). The Silver economy: brain training fired up by hard evidence. *Economist.* Available at: https://www.ft.com/content/c6028b80-3385-11e4-ba62-00144feabdc0

Corbett, A., Owen, A., Hampshire, A., Grahn, J., Stenton, R., Dajani, S., et al. (2015). The Effect of an online cognitive training package in healthy older adults: an online randomized controlled trial. *J. Am. Med. Dir. Assoc.* 16, 990–997. doi: 10.1016/j.jamda.2015.06.014

Daws, R. E., and Hampshire, A. (2017). The negative relationship between reasoning and religiosity is underpinned by a bias for intuitive responses specifically when intuition and logic are in conflict. *Front. Psychol.* 8:2191. doi: 10.3389/fpsyg.2017.02191

Edwards, J. D., Fausto, B. A., Tetlow, A. M., Corona, R. T., and Valdés, E. G. (2018). Systematic review and meta-analyses of useful field of view cognitive training. *Neurosci. Biobehav. Rev.* 84, 72–91. doi: 10.1016/j.neubiorev.2017.11.004

Engert, V., Kok, B. E., Papassotiriou, I., Chrousos, G. P., and Singer, T. (2017). Specific reduction in cortisol stress reactivity after social but not attention-based mental training. *Sci. Adv.* 3:e1700495. doi: 10.1126/sciadv.1700495

Ericsson, K. A. (2006). "The influence in experience and deliberate practice on the development of superior expert performance," in *The Cambridge Handbook of Expertise and Expert Performance*, eds K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman (Cambridge, MA: Cambridge University Press), 683–703.

Fissler, P., Küster, O. C., Loy, L. S., Laptinskaya, D., Rosenfelder, M. J., von Arnim, C. A. F., et al. (2017). Jigsaw puzzles as cognitive enrichment (PACE) - the effect of solving jigsaw puzzles on global visuospatial cognition in adults 50 years of age and older: study protocol for a randomized controlled trial. *Trials* 18:415. doi: 10.1186/s13063-017-2151-9

Foroughi, C. K., Monfort, S. S., Paczynski, M., McKnight, P. E., and Greenwood, P. M. (2016). Placebo effects in cognitive training. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7470–7474. doi: 10.1073/pnas.1601243113

Foy, J. G., and Mann, V. A. (2014). Adaptive cognitive training enhances executive control and visuospatial and verbal working memory in beginning readers. *Int. Educ. Res.* 2, 19–43.

Granic, I., Lobel, A., and Engels, R. C. (2014). The benefits of playing video games. *Am. Psychol.* 69, 66–78. doi: 10.1037/a0034857

Graziano, P. A., and Hart, K. (2016). Beyond behavior modification: benefits of social–emotional/self-regulation training for preschoolers with behavior problems. *J. Sch. Psychol.* 58, 91–111. doi: 10.1016/j.jsp.2016.07.004

Green, C. S., and Bavelier, D. (2015). Action video game training for cognitive enhancement. *Curr. Opin. Behav. Sci.* 4, 103–108.

Grossmann, I., Karasawa, M., Izumi, S., Na, J., Varnum, M. E. W., Kitayama, S., et al. (2012). Age and wisdom: culture matters. *Psychol. Sci.* 23, 1059–1066. doi: 10.1177/0956797612446025

Guye, S., and von Bastian, C. C. (2017). Working memory training in older adults: bayesian evidence supporting the absence of transfer. *Psychol. Aging* 32, 732–746. doi: 10.1037/pag0000206

Hampshire, A., Highfield, R. R., Parkin, B. L., and Owen, A. M. (2012). Fractionating human intelligence. *Neuron* 76, 1225–1237. doi: 10.1016/j.neuron.2012.06.022

Hardy, J. L., Nelson, R. A., Thomason, M. E., Sternberg, D. A., Katovich, K., Farzin, F., et al. (2015). Enhancing cognitive abilities with comprehensive training: a large. online, randomized, active-controlled trial. *PLoS One* 10:e0134467. doi: 10.1371/journal.pone.0134467

Holmes, J., Woolgar, F., Hampshire, A., and Gathercole, S. E. (2018). Are working memory training effects paradigm-specific? *Front. Psychol.* 10:1103. doi: 10.3389/fpsyg.2019.01103

Johnstone, S. J., Roodenrys, S., Blackman, R., Johnston, E., Loveday, K., Mantz, S., et al. (2012). Neurocognitive training for children with and without AD/HD. *Atten. Defic. Hyperact. Disord.* 4, 11–23. doi: 10.1007/s12402-011-0069-8

Karbach, J., and Verhaeghen, P. (2014). Making working memory work: a meta-analysis of executive-control and working memory training in older adults. *Psychol. Sci.* 25, 2027–2037. doi: 10.1177/0956797614548725

Katz, B. (2014). Brain-training isn't just a modern phenomenon, the edwardians were also fans. *Conversation* 1:2014.

Lampit, A., Hallock, H., Suo, C., Naismith, S. L., and Valenzuela, M. (2015a). Cognitive training-induced short-term functional and long-term structural plastic change is related to gains in global cognition in healthy older adults: a pilot study. *Front. Aging Neurosci.* 7:14. doi: 10.3389/fnagi.2015.00014

Lampit, A., Valenzuela, M., and Gates, N. J. (2015b). Computerized cognitive training is beneficial for older adults. *J. Am. Geriatr. Soc.* 63, 2610–2612.

Lindenberger, U., Wenger, E., and Lövdén, M. (2017). Towards a stronger science of human plasticity. *Nat. Rev. Neurosci.* 18, 261–262.

Mayas, J., Parmentier, F. B., Andrés, P., and Ballesteros, S. (2014). Plasticity of attentional functions in older adults after non-action video game training: a randomized controlled trial. *PLoS One* 9:e92269. doi: 10.1371/journal.pone.0092269

Melby-Lervåg, M., and Hulme, C. (2013). Is working memory training effective? a meta-analytic review. *Dev. Psychol.* 49, 270–291. doi: 10.1037/a0028228

Melby-Lervåg, M., and Hulme, C. (2016). There is no convincing evidence that working memory training is effective: a reply to Au et al. (2014) and Karbach and Verhaeghen (2014). *Psychon. Bull. Rev.* 23, 324–330. doi: 10.3758/s13423-015-0862-z

Mewborn, C. M., Lindbergh, C. A., and Stephen Miller, L. (2017). Cognitive interventions for cognitively healthy, mildly impaired, and mixed samples of older adults: a systematic review and meta-analysis of randomized-controlled trials. *Neuropsychol. Rev.* 27, 403–439. doi: 10.1007/s11065-017-9350-8

Moreau, D., and Conway, A. R. (2014). The case for an ecological approach to cognitive training. *Trends Cogn. Sci.* 18, 334–336. doi: 10.1016/j.tics.2014.03.009

Nouchi, R., Taki, Y., Takeuchi, H., Hashizume, H., Akitsuki, Y., Shigemune, Y., et al. (2012). Brain training game improves executive functions and processing speed in the elderly: a randomized controlled trial. *PLoS One* 7:e29676. doi: 10.1371/journal.pone.0029676

Nouchi, R., Taki, Y., Takeuchi, H., Hashizume, H., Nozawa, T., Kambara, T., et al. (2013). Brain training game boosts executive functions, working memory and processing speed in the young adults: a randomized controlled trial. *PLoS One* 8:e55518. doi: 10.1371/journal.pone.0055518

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., et al. (2010). Putting brain training to the test. *Nature* 465, 775–778.

Rebok, G. W., Ball, K., Guey, L. T., Jones, R. N., Kim, H. Y., King, J. W., et al. (2014). Ten-year effects of the advanced cognitive training for independent and vital elderly cognitive training trial on cognition and everyday functioning in older adults. *J. Am. Geriatr. Soc.* 62, 16–24. doi: 10.1111/jgs.12607

Rohwedder, S., and Willis, R. J. (2010). Mental retirement. *J. Econ. Perspec.* 24, 119–138.

Sala, G., Tatlidil, K. S., and Gobet, F. (2018). Video game training does not enhance cognitive ability: a comprehensive meta-analytic investigation. *Psychol. Bull.* 144, 111–139. doi: 10.1037/bul0000139

Seitz, A. R. (2017). A new framework of design and continuous evaluation to improve brain training. *J. Cogn. Enhance* 2, 78–87. doi: 10.1007/s41465-017-0058-8

SharpBrains (2013). *Executive Summary: Infographic on the Digital Brain Health Market 2012-2020*. Available at: http://sharpbrains.com/executive-summary (accessed March 4, 2017).

Shimamura, A. P., Berry, J. M., Mangels, J. A., Rusting, C. L., and Jurica, P. J. (1995). Memory and cognitive abilities in university professors: evidence for successful aging. *Psychol. Sci.* 6, 271–277.

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., et al. (2016). Do "Brain-Training". programs work? *Psychol. Sci. Public Interest* 17, 103–186.

Simonton, D. K. (1990). "Creativity and wisdom in aging," in *Handbook of the Psychology of Aging*, 3rd Edn, eds J. E. Birren and K. W. Schaie (New York, NY: Academic Press), 320–329.

Smith, G. E., Housen, P., Yaffe, K., Ruff, R., Kennison, R. F., Mahncke, H. W., et al. (2009). A cognitive training program based on principles of brain plasticity: results from the improvement in memory with plasticity-based adaptive cognitive training (IMPACT) study. *J. Am. Geriatr. Soc.* 57, 594–603. doi: 10.1111/j.1532-5415.2008.02167.x

Soveri, A., Antfolk, J., Karlsson, L., Salo, B., and Laine, M. (2017). Working memory training revisited: a multi-level meta-analysis of n-back training studies. *Psychon. Bull. Rev.* 24, 1077–1096. doi: 10.3758/s13423-016-1217-0

Staudinger, U. M., and Baltes, P. B. (1996). Interactive minds: a facilitative setting for wisdom-related performance? *J. Pers. Soc. Psychol.* 71, 746–762.

Steiner, N. J., Sheldrick, R. C., Gotthelf, D., and Perrin, E. C. (2011). Computer-based attention training in the schools for children with attention deficit/hyperactivity disorder: a preliminary trial. *Clin. Pediatr.* 50, 615–622. doi: 10.1177/0009922810397887

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *J. Int. Neuropsychol. Soc.* 8, 448–460.

Strobach, T., and Huestegge, L. (2017). Evaluating the effectiveness of commercial brain game training with working-memory tasks. *J. Cogn. Enhance* 1, 539–558.

Thorell, L. B., Lindqvist, S., Bergman Nutley, S., Bohlin, G., and Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Dev. Sci.* 12, 106–113. doi: 10.1111/j.1467-7687.2008.00745.x

Toril, P., Reales, J. M., and Ballesteros, S. (2014). Video game training enhances cognition of older adults: a meta-analytic study. *Psychol. Aging.* 29, 706–716. doi: 10.1037/a0037507

Unsworth, N., Redick, T. S., McMillan, B. D., Hambrick, D. Z., Kane, M. J., and Engle, R. W. (2015). Is playing video games related to cognitive abilities? *Psychol. Sci.* 26, 759–774. doi: 10.1177/0956797615570367

Valk, S. L., Bernhardt, B. C., Trautwein, F. M., Böckler, A., Kanske, P., Guizard, N., et al. (2017). Structural plasticity of the social brain: differential change after socio-affective and cognitive mental training. *Sci. Adv.* 3:e1700489. doi: 10.1126/sciadv.1700489

Vermeij, A., Claassen, J. A., Dautzenberg, P. L., and Kessels, R. P. (2016). Transfer and maintenance effects of online working-memory training in normal ageing and mild cognitive impairment. *Neuropsychol. Rehabil.* 26, 783–809. doi: 10.1080/09602011.2015.1048694

Wang, P., Liu, H. H., Zhu, X. T., Meng, T., Li, H. J., and Zuo, X. N. (2016). Action video game training for healthy adults: a meta-analytic study. *Front. Psychol.* 7:907. doi: 10.3389/fpsyg.2016.00907

Willis, S. L., Tennstedt, S. L., Marsiske, M., Ball, K., Elias, J., Koepke, K. M., et al. (2006). Long-term effects of cognitive training on everyday functional outcomes in older adults. *JAMA* 296, 2805–2814.

Wolinsky, F. D., Vander Weg, M. W., Howren, M. B., Jones, M. P., and Dotson, M. M. (2013). A randomized controlled trial of cognitive training using a visual speed of processing intervention in middle aged and older adults. *PLoS One* 8:e61624. doi: 10.1371/journal.pone.0061624

Wolinsky, F. D., Vander Weg, M. W., Howren, M. B., Jones, M. P., and Dotson, M. M. (2016). Effects of cognitive speed of processing training on a composite neuropsychological outcome: results at one-year from the IHAMS randomized controlled trial. *Int. Psychogeriatr.* 28, 317–330. doi: 10.1017/S1041610215001428

Zelinski, E. M., Spina, L. M., Yaffe, K., Ruff, R., Kennison, R. F., Mahncke, H. W., et al. (2011). Improvement in memory with plasticity-based adaptive cognitive training: results of the 3-month follow-up. *J. Am. Geriatr. Soc.* 59, 258–265. doi: 10.1111/j.1532-5415.2010.03277.x

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

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 09 August 2019
doi: 10.3389/fnhum.2019.00275

Differences in Brain Activity After Learning With the Use of a Digital Pen vs. an Ink Pen—An Electroencephalography Study

Kiyoyuki Osugi[1,2†], Aya S. Ihara[1†], Kae Nakajima[1], Akiyuki Kake[3], Kizuku Ishimaru[3], Yusuke Yokota[1] and Yasushi Naruse[1]*

[1]Center for Information and Neural Networks, National Institute of Information and Communications Technology, Osaka University, Kobe, Japan, [2]Graduate School of Frontier Bioscience, Osaka University, Suita, Japan, [3]Wacom Co., Ltd., Kazo, Japan

The purpose of this study is to clarify whether there is a learning effect on brain activity after writing with an ink pen vs. a digital pen. Previous studies have reported the superiority of handwriting to typing in terms of learning performance, but differences between the use of an ink pen vs. a digital pen remain unclear. In the present study, the participants learned to read difficult words by writing with an ink pen vs. a digital pen. After the learning period, electroencephalography (EEG) signals were measured, while the participants underwent a repetition priming paradigm with the use of the learned words. The repetition priming effect of the N400 event-related potential (ERP) was quantified as an index of the learning effect and the effects between pen types were compared. The groups were also subdivided according to whether a digital pen is frequently used (familiar vs. unfamiliar group). The number of writing repetitions for each word within 10 min during the learning activity and the post-learning test scores were not affected by the pen-type or familiarity with a digital pen. However, the repetition priming effect of the N400 was greater for words written with a digital pen in the learning session, as compared with an ink pen, in the familiar group, but not the unfamiliar group. These results suggest that for those familiar with its use, writing with a digital pen may improve learning relative to the use of an ink pen.

Keywords: digital device, learning, electroencephalography, digital pen, handwriting

OPEN ACCESS

Edited by:
Carmen Moret-Tatay, Catholic University of Valencia San Vicente Mártir, Spain

Reviewed by:
Rui Alexandre Alves, University of Porto, Portugal
Jean-luc Velay, UMR7291 Laboratoire de Neurosciences Cognitives (LNC), France

***Correspondence:**
Aya S. Ihara
ihara@nict.go.jp

†These authors have contributed equally to this work

Received: *29 March 2019*
Accepted: *23 July 2019*
Published: *09 August 2019*

Citation:
Osugi K, Ihara AS, Nakajima K, Kake A, Ishimaru K, Yokota Y and Naruse Y (2019) Differences in Brain Activity After Learning With the Use of a Digital Pen vs. an Ink Pen—An Electroencephalography Study. Front. Hum. Neurosci. 13:275. doi: 10.3389/fnhum.2019.00275

INTRODUCTION

Digital devices are increasingly used in education, so understanding the differences in the effects on learning ability relative to the use of analog devices could lead to more effective educational practices. Experimental psychological studies have investigated the differences in learning effects between writing with a conventional pen and typing on a keyboard. In adult participants, recognition accuracy was higher after they learned unfamiliar characters by writing them down on paper than typing on a keyboard (Longcamp et al., 2008), and more words were recalled after writing on paper than typing (Mangen et al., 2015). Preschool children also learned letters and words more effectively by handwriting than typing (Longcamp et al., 2005; Kiefer et al., 2015). Additionally, handwriting seems to be more effective for conceptual comprehension than typing.

In fact, comprehension assessment of listening to technology/entertainment/design talks was superior among college students who made notes in a notebook using a pen as compared to those who typed notes on a laptop computer (Mueller and Oppenheimer, 2014). The advantage of handwriting over typing has also been indicated in neuroscientific approaches using electroencephalography (EEG; van der Meer and van der Weel, 2017) and magnetic resonance imaging (Vinci-booher et al., 2016).

However, few studies have investigated whether learning effects differ between writing with a digital pen on a tablet vs. writing with a conventional pen on paper. Hatano et al. (2015) conducted an EEG experiment in which the participants took notes with a digital pen on a tablet or with a mechanical pencil on paper while listening to scientific lessons. There were no significant differences in the scores of comprehension and memory tests performed after taking notes on a tablet vs. paper. However, theta-band (4–7 Hz) EEG activity was higher when writing on a tablet than on paper. Because the theta-band EEG activity was reported to increase with the cognitive load (Gevins et al., 1997; Borghini et al., 2012; Anguera et al., 2014), the use of a digital pen and tablet require more cognitive effort to monitor written characters and more attention to movements for writing compared with writing on paper. In fact, recent studies reported that the movements of handwriting with a digital pen on a tablet are not the same as with a conventional pen on paper (Alamargot and Morin, 2015; Gerth et al., 2016a,b; Wollscheid et al., 2016; Guilbert et al., 2019). Alamargot and Morin (2015) demonstrated that the handwriting kinematics of school-age children when writing with a plastic-tipped pen on a tablet screen is different from when writing with a ballpoint pen on paper, and the effects vary with age, as second graders made longer pauses and nine graders increased pen pressure and speed. These results suggest that segment trajectory calculation is disturbed in younger children and control of muscular adjustment is disturbed in older children. From the points of view of movements and brain activities, handwriting with a digital pen on a tablet might disturb cognitive activities, such as learning. To the best of our knowledge, this is the first study to investigate whether brain activity after learning by handwriting with a digital pen on a tablet is different from that with a pen on paper. The difference of after-effect might vary according to the (un)familiarity with a digital device.

In the present study, the N400, an event-related potential (ERP) response, was used to measure the learning effect of handwriting with a digital pen vs. an ink pen. The N400 is a negative-going component peaking around 400 ms after exposure to words, pictures, and other meaningful stimuli (for a review, see Kutas and Federmeier, 2011), which is related to semantic processing so that the amplitude changes with the ease of accessing information from long-term memory and integrating semantic representations into a preceding context (for a review, see Kutas and Federmeier, 2000). Many studies have shown that the N400 changes with language learning (Ojima et al., 2005, 2011) and developmental progress (Friedrich and Friederici, 2004, 2010; Reid et al., 2009). Most importantly, the N400 effects have been observed in an earlier stage of learning than behavioral indices (McLaughlin et al., 2004). Regarding an adult's ability to learn a second language (L2), McLaughlin et al. (2004) showed that the amplitude modulation of N400 discriminated between L2 words and pseudo-words after 14 h of classroom instruction, while the participants reached only chance levels when making overt L2 word-nonword judgments, suggesting that the N400 is a powerful tool to reveal the effects of learning, especially in the early stage.

The Japanese language has two writing systems [i.e., kana (syllabograms) and kanji (morphograms)], and so, many words have two notations. As the learning content, well-known Japanese words are used that are generally written in kana (syllabograms), as most Japanese people cannot read. Although there are two notations, most are familiar with the kana notation, but not kanji notation (**Figure 1**). In the learning activity, the participants learned the readings of such words by writing with a digital pen on a tablet and with an ink pen on paper. Just after the learning activity, EEG experiments were conducted with a repetition priming paradigm to determine whether the repetition priming effects of the N400 were affected by learning tools (i.e., digital pen vs. ink pen) and/or familiarity with a digital pen and tablet. When two stimuli are presented consecutively and the subsequent stimulus (target) is identical/related to the preceding stimulus (prime), the N400 amplitude of the target decreases relative to the unrepeated/unrelated target (van Petten et al., 1991; Deacon et al., 2004; Matsumoto et al., 2005; Rugg, 1985; Holcomb, 1993). In the present study, the words written in the learning activity (prime) were followed by kana words (target) that were semantically and phonologically identical to the prime words (repetitive condition) or not (non-repetitive condition). We assumed that as learning progresses, a larger difference in the N400 amplitude between the repetitive and non-repetitive conditions (i.e., repetition priming effect of the N400) would occur. The differences in the repetition priming effect of the N400 with writing with a digital pen on a tablet vs. with an ink pen on paper were used for comparisons between participants familiar and unfamiliar with the digital pen system.

MATERIALS AND METHODS

Participants

Twenty-eight healthy volunteers participated in the study. They were divided into two groups according to the results of a questionnaire distributed after the EEG experiment: 11 participants (10 men and one woman; age, 26–52 years) who used a digital pen/tablet system in their daily lives (familiar group) and 17 participants (10 men and seven women; age, 21–47 years) who did not (unfamiliar group). The study protocol was approved by the Bioinformatics Ethics Committee of the National Institute of Information and Communications Technology, and all participants provided informed written consent before participation in this study.

Learning Materials

Selection of Words for the Learning Activity

As the learning contents, 120 well-known Japanese words were selected that are generally written in kana (syllabograms), as

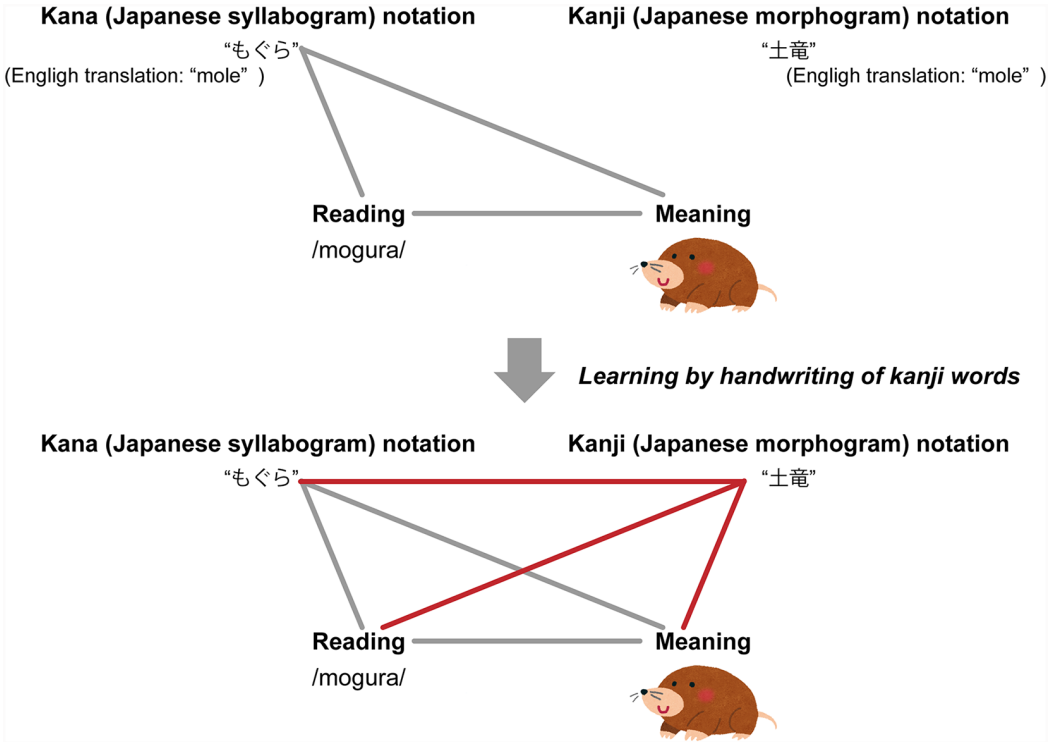


FIGURE 1 | An example of a Japanese word written in kanji that is difficult to read. Mole in English corresponds to two notations in Japanese: "もぐら" written in kana (syllabogram) and "土竜" written in kanji (morphogram). However, the kana notation is generally used and most Japanese cannot read the kanji notation, so only the kana notation was associated with the reading and meaning of the word "mole," but not the kanji notation. We assumed that in the learning activity, the participants wrote the kanji notations of such words, then the kanji notations have association with the reading and meaning, as well as the kana notation.

most Japanese people cannot read written kanji (morphograms). The words were primary school level that are familiar and easy to image for Japanese people. Actually, the words had high familiarity and imaginability values of >5.7 on average on a seven-point scale (Amano and Kondo, 1999; Sakuma et al., 2005; **Table 1**). The 120 words were divided into six sets of 20 words each. We confirmed that the lexical properties of the words were matched among the sets based on the results of the Kruskal–Wallis test, where there were no significant differences in the number of characters (p = 0.20), number of morae (p = 0.98), familiarity values (p = 0.50), and imaginability values (p = 0.73) across the sets. In the learning activity, each participant learned two sets of words that were randomly selected across participants.

Stimuli for EEG Experiment

Forty kanji words that each participant wrote in the learning activity (i.e., 20 with an ink pen and 20 with a digital pen) were used as the prime stimuli in the repetition priming paradigm. Each prime stimulus was followed by the target stimuli, which comprised words written in Japanese syllabograms (kana). According to the type of target word, two conditions were set up: repetitive, where the target stimulus represented the reading of the prime stimulus (i.e., semantically and phonologically identical), and non-repetitive, where the target stimulus did not represent the reading of the prime stimulus (i.e., semantically and phonologically different). Trials under the non-repetitive condition used kanji words from one set and readings from the other sets (e.g., pair of kanji words of set A and readings of set B). To prevent unwanted influences from phonological and semantic priming effects on the non-repetition condition, two evaluators checked the presence of phonological similarity and semantic relationships between the prime and target words in each pair. Finally, the word pairs that the both evaluators judged as having no phonological similarity or semantic relationships were adopted. Each kanji word

TABLE 1 | Lexical properties of word sets.

Lexical property	Set A	Set B	Set C	Set D	Set E	Set F	Kruskal–Wallis test
Number of characters	2.2 ± 0.5	2.1 ± 0.4	2.2 ± 0.6	2.0 ± 0.5	2.3 ± 0.5	2.0 ± 0.5	p = 0.20
Numbers of morae	3.5 ± 0.5	3.5 ± 0.5	3.5 ± 0.5	3.5 ± 0.5	3.4 ± 0.5	3.4 ± 0.6	p = 0.98
Familiarity values	5.8 ± 0.5	5.7 ± 0.9	6.0 ± 0.5	5.9 ± 0.6	5.8 ± 0.7	6.0 ± 0.5	p = 0.50
Imaginability values	5.8 ± 0.6	5.7 ± 0.7	5.9 ± 1.0	5.7 ± 0.7	5.8 ± 0.7	5.7 ± 0.6	p = 0.73

learned by each participant were presented four times as the prime stimulus throughout the experiment: two times for the repetitive condition and two times for the non-repetitive condition. Hence, the participants underwent 160 trials in total: 40 trials with the repetitive condition and 40 trials with the non-repetitive condition in which the prime stimuli were the words learned with a digital pen, as well as 40 trials with the repetitive condition and 40 trials with the non-repetitive condition in which the prime stimuli were the words learned with an ink pen.

Experimental Procedures

The experiment flow was as follows: (1) pre-learning test; (2) learning activity; (3) post-learning test; (4) EEG measurement; and (5) questionnaire survey (**Figure 2**). The protocols for each of these steps are described below.

Pre- and Post-learning Tests

To investigate the effects of the learning activity on test performance, the participant's ability to read the selected kanji words was tested both before and after learning. The test sheet given to the participants contained all of the kanji words and the participants answered by writing the correct reading of the words in kana. The words already known by the participant before the learning session in the pre-learning test (**Supplementary Figure S1**) were excluded from further analyses of both performance and the N400.

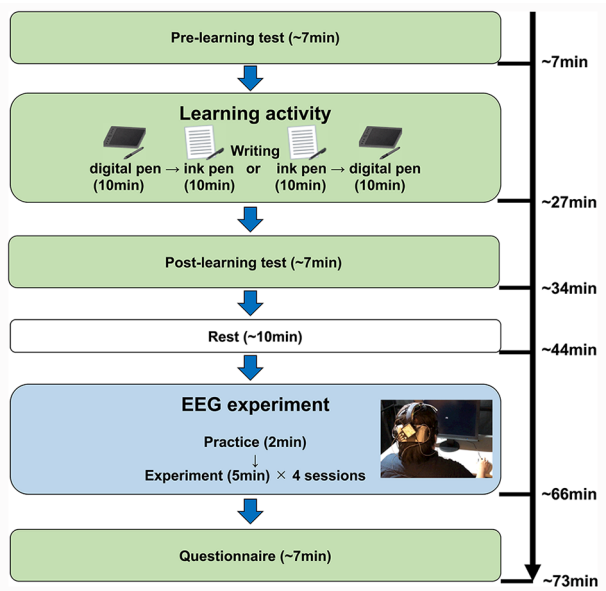


FIGURE 2 | Experimental steps. This chart shows the detailed order of the experimental steps. First, the participants completed a pre-learning test of about 7 min. For the learning activity, the participants learned the kanji notations by writing with each device (10 min each). The order of device use was randomized. Then, the participants completed the post-learning test. After 10-min rest, the electroencephalography (EEG) experiment was conducted for about 22 min. Finally, they answered the questionnaire (about 7 min).

Learning Activity

For each participant, two sessions were conducted in the learning activity: a digital pen session, during which the participants wrote 20 words of one set repeatedly with a digital pen on a tablet, and an ink pen session, during which they wrote 20 words of the other set with an ink pen on papers. Each session with a given pen-type lasted 10 min. The two sessions were performed in random order among the participants. There was a rest period between the two sessions.

Twenty pairs of kanji words and the corresponding readings (written with kana syllabograms) were written on each learning sheet. The participants were asked to copy the kanji words and to memorize the readings and were told that the readings after the learning activity would be tested.

For the digital pen learning session, a PDF file of the learning sheet was displayed to the participants on a tablet (Cintiq 13HD Creative Pen Display DTK-1301; Wacom Co., Limited, Tokyo, Japan), while the participants wrote with a digital pen (Propen; Wacom). For the ink pen session, the learning sheet was physically placed on a tablet (Intuos Pro Large PTH-851; Wacom) and the participants were instructed to write with an ink pen (Wacom).

EEG Measurement

The experiment described in "Stimuli for EEG Experiment" section was conducted while recording the EEG signals. Briefly, the prime, target, and cue (###) were presented continuously with each stimulus-onset asynchrony set to 1,000 ms (**Figure 3**). The presentation duration for the prime and target stimuli was 300 ms, while that of the cue was 500 ms. The participant was asked to read the prime (kanji) and target (kana) words silently and to answer whether the readings of the prime and target stimuli matched or not by clicking the computer mouse after presentation of the cue. The prime for the next trial was presented 2,000–3,000 ms after the cue onset.

EEG and electrooculography (EOG) signals were measured continuously using an eight-channel wearable EEG device (Polimate mini AP108; Miyuki Giken Co. Limited, Tokyo, Japan). The dry midline electrodes (Unique Medical Co. Limited, Tokyo, Japan) Fz, Cz, and Pz were used according to the International 10-20 system. In addition, to detect the artifacts of eye movements and blinks, and to reject the noise components from the EEG signals, an electrode was placed on the upper and right sides of the left eye to measure the vertical and horizontal EOG components. All signals were sampled at 500 Hz with the use of the left earlobe as the ground and the right earlobe as the reference.

Questionnaire Survey

After completing the EEG measurements, the participants responded to questionnaires concerning the use of digital pens on a daily basis as well as the familiarity with the devices used in the learning activity. Each participant was asked to choose either the digital pen or the ink pen with respect to: (1) which they felt required a greater workload; (2) was more enjoyable; (3) easier to memorize; and (4) easier to write with.

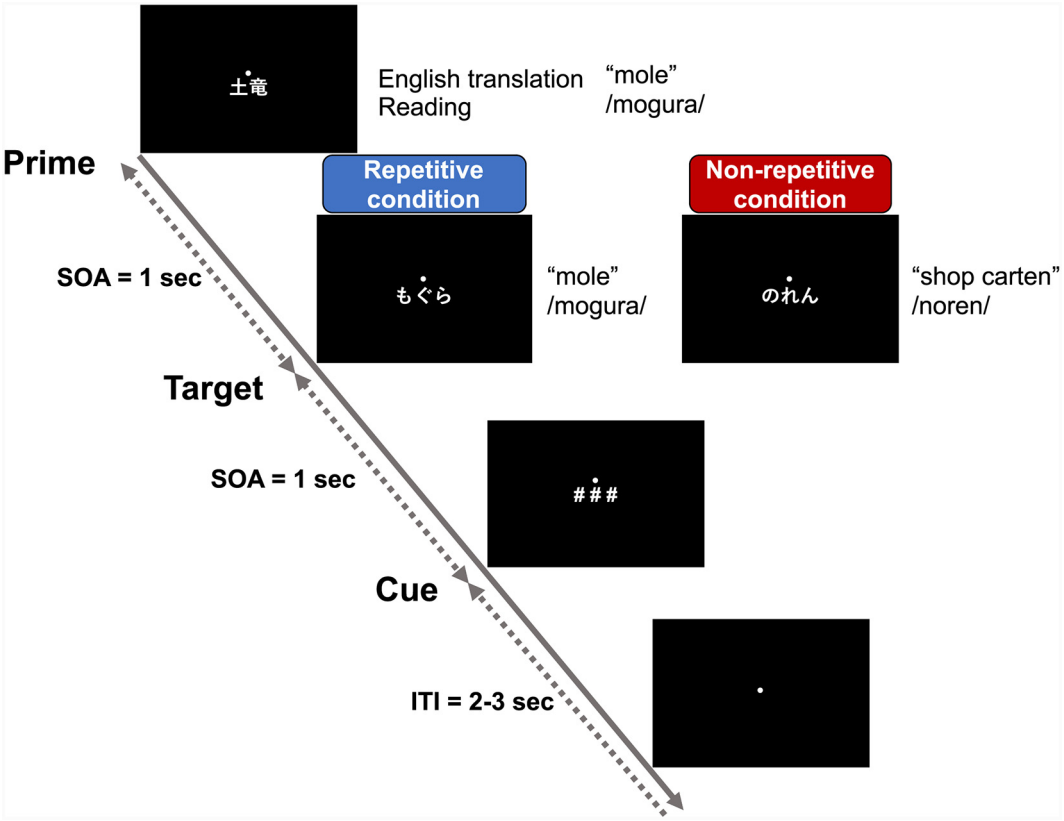


FIGURE 3 | Schematic representation of the repetition priming paradigm. The prime stimuli were words written in morphograms (kanji), which the participants wrote in the learning activity. The target stimuli were words written in syllabograms (kana). In the repetitive condition, the target words were semantically and phonologically identical to the prime stimuli, but distinctly different in the non-repetitive condition. After presentation of the cue (###), the participants answered whether the reading of the target was correct or not by clicking a mouse. The prime, target, and cue were presented with a stimulus-onset asynchrony of 1 s. The intertrial interval between the offset of the cue and the onset of the next prime was randomly set at 2–3 s.

Data Analyses

Performance

The difference between the number of correct answers on the reading test before and after the learning activity was recorded as the number of words memorized. To assess the differences between groups and devices, two-factor, mixed-design analysis of variance was performed with factors of familiarity (familiar vs. unfamiliar group) and learning device (digital vs. ink pen).

N400

Analysis of the EEG and EOG signals was conducted using MATLAB (MathWorks Inc., Natick, MA, USA) and the EEGLAB toolbox[1]. A bandpass filter of 0.2–30 Hz was applied to the measured EEG and EOG signals, and artifact components, mainly caused by eye movements and blinks, were excluded from the EEG signals using noise reduction processing with artifact subspace reconstruction and independent component analysis. Next, the signals from 100 ms before to 800 ms after target onset were averaged for each condition (repetitive vs. non-repetitive) and each channel, and corrected using 100 ms before target onset as a baseline. Trials exceeding ±80 μV on the Fz, Cz, and Pz channels, those exceeding ±100 μV on the vertical and horizontal EOG channels, and those of words correctly adjusted based on preliminary reading tests were excluded from the average.

The Cz or Pz electrode with a large repetition priming effect on N400 was used as an analytical target for each participant and each learning device. The recorded N400 waves had an average amplitude of 300–450 ms after target onset.

Effects of familiarity and learning device on the repetition priming effect of the N400 were analyzed using the same two-factor analysis of variance described in “Performance” section. Furthermore, as a *post hoc* test, the *t*-test was conducted within each group to identify differences in the repetition priming effect between the two devices.

Questionnaire Survey

The χ^2 test was performed to determine whether the familiar group differed from the unfamiliar group regarding the proportion of participants who responded that workload, enjoyability, ease of memorization, and ease of writing were greater with the use of a digital pen vs. an ink pen.

[1]https://sccn.ucsd.edu/eeglab/index.php

RESULTS

Learning Activity and Pre/Post-learning Tests

Regarding the number of writing repetitions per word, there was no main effect of the participant group ($F_{(1,26)} = 0.78$; $p = 0.39$, partial $\eta^2 = 0.03$), main effect of the learning device ($F_{(1,26)} = 0.04$; $p = 0.84$, partial $\eta^2 = 0.00$), or interaction between the two factors ($F_{(1,26)} = 2.98$; $p = 0.10$, partial $\eta^2 = 0.10$; **Figure 4A**). Similarly, with regard to the number of words memorized, there was no main effect of the participant group ($F_{(1,26)} = 0.03$; $p = 0.87$, partial $\eta^2 = 0.00$), main effect of the learning device ($F_{(1,26)} = 2.94$; $p = 0.10$, partial $\eta^2 = 0.10$), or interaction ($F_{(1,26)} = 0.50$; $p = 0.49$, partial $\eta^2 = 0.02$; **Figure 4B**). Thus, performance in learning was not affected by either familiarity or the learning device used.

EEG Experiment

Regarding the accuracy rates for the task during the EEG experiment, there was no main effect of the participant group ($F_{(1,26)} = 0.10$; $p = 0.76$, partial $\eta^2 = 0.00$) or learning device ($F_{(1,26)} = 0.02$; $p = 0.90$, partial $\eta^2 = 0.00$) and no interaction between the participant group and learning device ($F_{(1,26)} = 0.02$; $p = 0.90$, partial $\eta^2 = 0.00$; **Figure 5A**).

For each participant group and device, the N400 (ERP; 350–450 ms) with the repetitive condition was smaller than with the non-repetitive condition (**Figure 6A**). With specific regard to the repetition priming effect of the N400 (i.e., difference between N400 amplitudes measured under non-repetitive and repetitive conditions; **Figure 6B**), a main effect of the learning device was observed ($F_{(1,26)} = 4.78$; $p = 0.04$, partial $\eta^2 = 0.16$). That is, the repetition priming effect was greater with the use of a digital

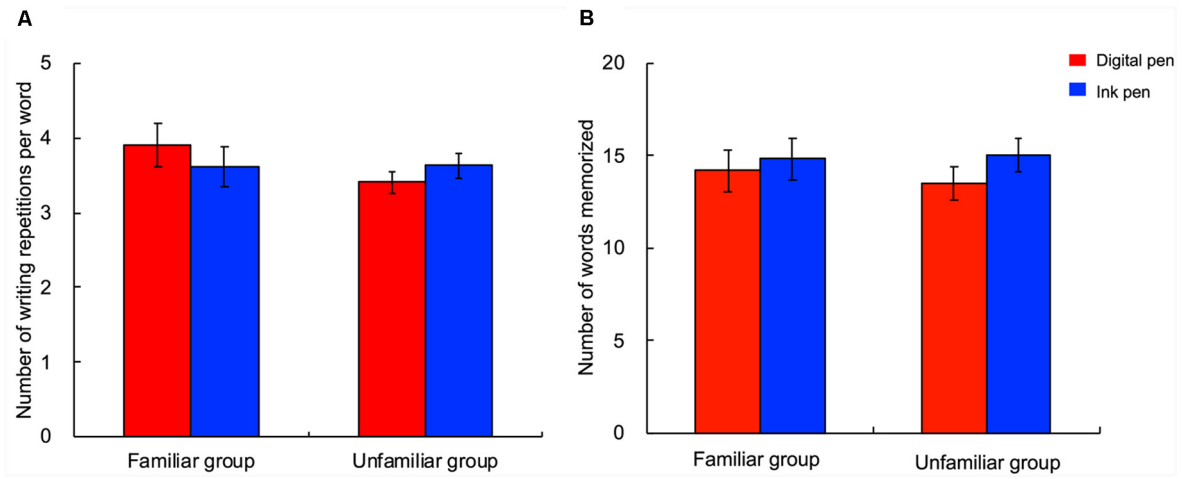


FIGURE 4 | Number of writing repetitions per word and number of words memorized. The number of writing repetitions per word during the learning activity **(A)** and the number of memorized words **(B)** were not affected by either the writing device or familiarity with the digital device. Each bar shows the grand average of the participants. The error bar represents the standard error. The red and blue bars indicate the use of a digital and ink pen, respectively.

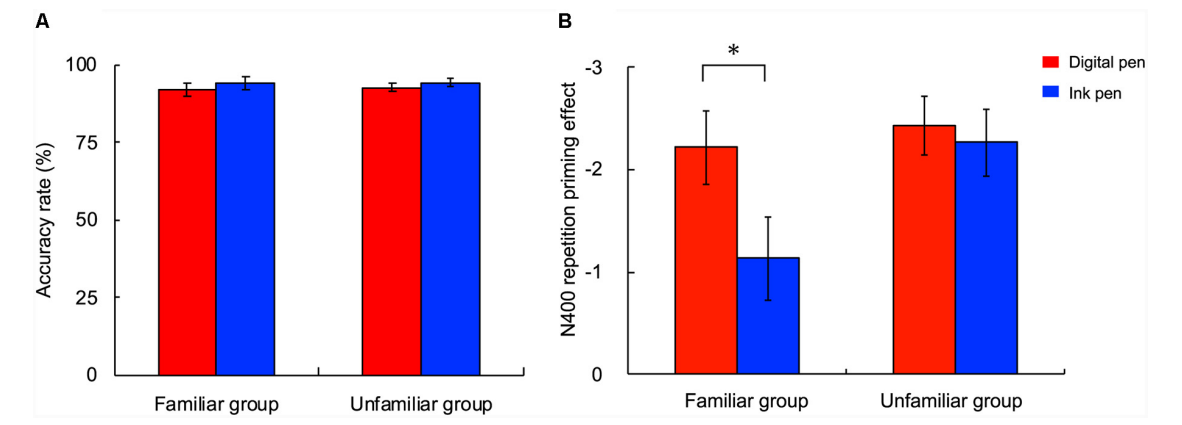


FIGURE 5 | Accuracy rate and repetition priming effect of the N400 in the EEG experiment. **(A)** The accuracy rate was not affected by the learning device or group. **(B)** The repetition effect of the N400 was significantly greater for words learned with the digital pen (red) than with the ink pen (blue) in the familiar group, but not in the unfamiliar group. The bars indicate mean amplitudes of the difference-event-related potential (ERP) from 300 to 450 ms (i.e., non-repetitive condition minus repetitive condition) between groups. The error bars represent the standard error. *$p < 0.05$.

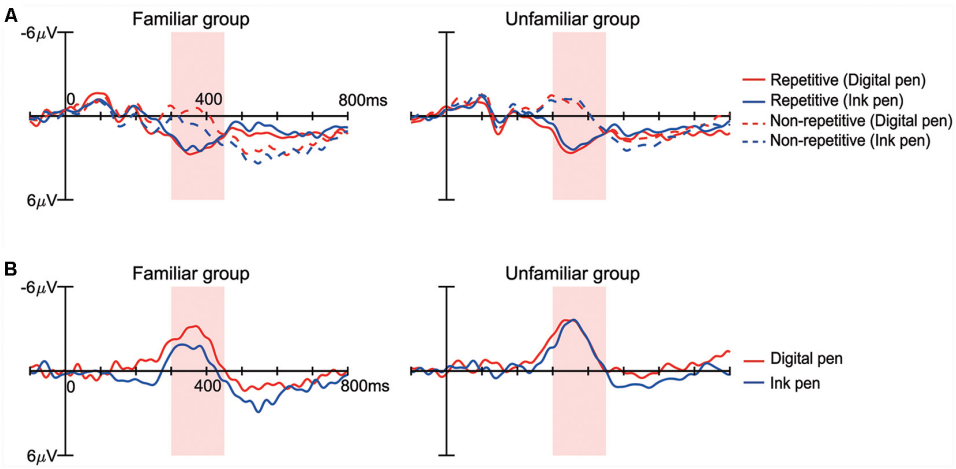


FIGURE 6 | ERPs measured in the repetition priming paradigm. **(A)** The ERPs from approximately 300–450 ms varied between the repetitive condition (solid line) and non-repetitive condition (dashed line) in both the familiar and unfamiliar groups. The red and blue lines indicate the digital and ink pen condition, respectively. **(B)** The difference in waveforms obtained by subtracting the repetitive condition from the non-repetitive condition had a negative peak at approximately 300–450 ms. In the familiar group, the amplitude was smaller for the ink pen (blue) than the digital pen (red), while there was no difference in the unfamiliar group.

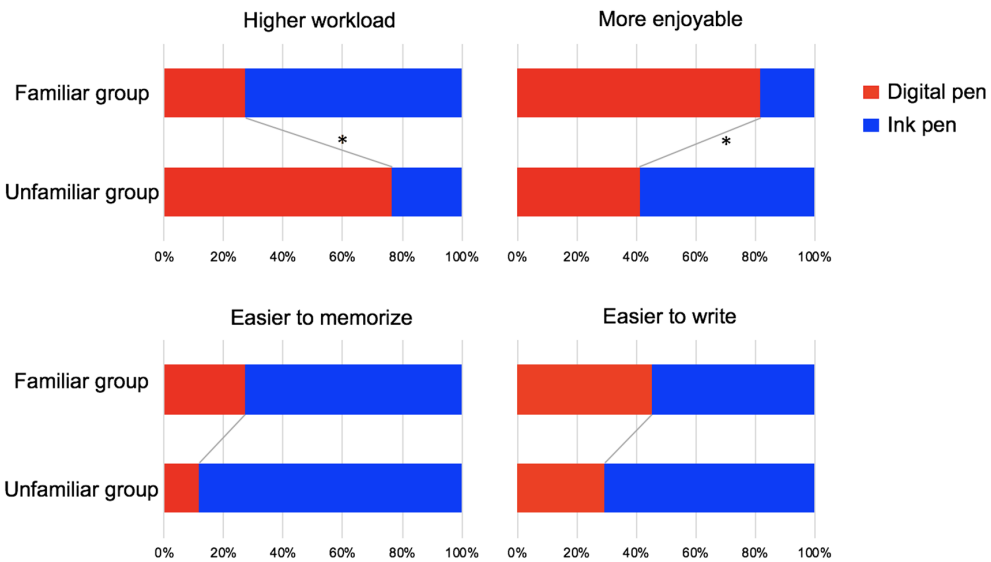


FIGURE 7 | Results of the questionnaire survey. Regarding the number of participants who answered "digital pen" (red) and "ink pen" (blue) to each survey item, the proportions differed significantly between the familiar and unfamiliar groups with respect to the degree of workload ($p = 0.01$) and enjoyability ($p = 0.03$). A greater number of participants in the familiar group (8 of 11) answered that "the use of the ink pen required a greater workload," whereas most participants in the unfamiliar group (13 of 17) answered that "the use of the digital pen required a greater workload." In addition, most participants in the familiar group (9 of 11) answered that "the digital pen was more enjoyable to use," whereas a slight majority of participants in the unfamiliar group (10 of 17) answered that "the ink pen was more enjoyable to use." There was no significant difference between groups regarding memorability and ease of writing. $^*p < 0.05$.

pen vs. an ink pen. In contrast, there was no main effect of the participant group ($F_{(1,26)} = 2.81$; $p = 0.11$, partial $\eta^2 = 0.10$) or interaction between the participant group and learning device ($F_{(1,26)} = 2.58$; $p = 0.12$, partial $\eta^2 = 0.09$).

When the differences between the learning devices were examined within each participant group as a *post hoc* test (**Figure 5B**), the repetition priming effect was significantly greater in the familiar group with the use of a digital pen vs. an ink pen ($t_{(10)} = -2.37$; $p = 0.02$, $d = 0.90$), while the effect was similar between pen types in the unfamiliar group ($t_{(16)} = -0.47$; $p = 0.32$, $d = 0.13$).

To determine whether the difference in the N400 effect between the familiar and unfamiliar groups resulted from differences in the ratio of men to women (i.e., familiar group: 10 men and one woman; unfamiliar group: 10 men and seven women), the data of only men were also analyzed because

the number of men was the same in the two groups. The results showed that the same effect was obtained from all participants; that is, the familiar group had a greater priming effect of the N400 for the words learned with a digital pen than those with an ink pen, while in the unfamiliar group, there was no effect by the learning device (**Supplementary Figure S2**). Therefore, the difference obtained among all participants likely did not result from differences in the sex ratio.

Questionnaire Survey

Figure 7 displays the tallied results of the participants' answers to the survey questions concerning workload, enjoyability, ease of memorization, and ease of writing. Regarding the number of participants who answered "digital pen" and "ink pen" to each survey item, the proportions differed significantly between the familiar and unfamiliar groups with respect to the degree of workload ($\chi^2 = 6.60$; $p = 0.01$, $\varphi = 0.49$) and enjoyability ($\chi^2 = 4.50$; $p = 0.03$, $\varphi = 0.40$). In the familiar group, a larger number of participants (8 of 11) than in the unfamiliar group answered that "the use of the ink pen required a greater workload." In the unfamiliar group, most participants (13 of 17) answered that "the use of the digital pen required a greater workload." In addition, most participants in the familiar group (9 of 11) responded that "the digital pen was more enjoyable to use." Conversely, in the unfamiliar group, a slight majority of participants (10 of 17) responded that "the ink pen was more enjoyable to use." Meanwhile, no significant differences were observed between the two participant groups with respect to memorability and ease of writing.

DISCUSSION

In the present study, the modulation of electrophysiological signals was investigated by writing words with a digital pen vs. a conventional ink pen, while comparing between participants familiar vs. unfamiliar with the digital device. Regarding the performance level, there were no differences in the number of writing repetitions for each word over a 10-min period and the number of words memorized afterward between the learning devices in both groups, or in the accuracy rate of the EEG experiment. In contrast, the repetition priming effect of the N400 detected immediately after learning showed the difference between the learning devices. In the familiar group, the repetition priming effect of the N400 was significantly greater for the words that were learned by writing with a digital pen vs. an ink pen, but not in the unfamiliar group. The dissociation of performance and the N400 effect might be due to differences in the sensitivity to detect learning effects. That is, the N400 effect was more sensitive to early learning compared with the behavioral assessment (McLaughlin et al., 2004). Therefore, our result indicates that when the participants are familiar with the use of digital pens, learning with the ink pen gave rise to lesser effect on brain activity.

From these results, the question arises why the use of a digital device might affect brain activity. Our questionnaire survey indicated that more participants in the familiar group found it fun to use the digital pen and felt less workload when using the device, as compared with the participants in the unfamiliar group. Previous studies have reported that mood can affect learning (Nadler et al., 2010; Bakic et al., 2014) and language processing (Federmeier et al., 2001; Vissers et al., 2010, 2013; Chwilla et al., 2011). Therefore, the difference in the N400 effect between the learning devices might be caused by differences in mood while using a particular device (i.e., less enjoyable and higher workload when writing with an ink pen). However, the data obtained in this study were insufficient for further investigation, which is a limitation to this study. Nonetheless, further research on this topic is warranted.

DATA AVAILABILITY

All datasets generated for this study are included in the manuscript and/or the **Supplementary Files**.

ETHICS STATEMENT

The study was conducted after approval was obtained from the Bioinformatics Ethics Committee of the National Institute of Information and Communications Technology, and all participants provided informed written consent in advance regarding their participation in this study.

AUTHOR CONTRIBUTIONS

AI, YN, KN, YY, and AK designed the experiment. KN prepared the materials and experimental equipment. KN, KO, and KI conducted the experiments. KN and KO analyzed the data. KO and AI wrote the first draft of the manuscript, and all authors revised the draft and approved the final version.

ACKNOWLEDGMENTS

We wish to thank Mrs. Kazuko Otani for her assistance in preparing word lists and helping with data acquisition.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2019.00275/full#supplementary-material

REFERENCES

Alamargot, D., and Morin, M. F. (2015). Does handwriting on a tablet screen affect students' graphomotor execution? A comparison between Grades Two and Nine. *Hum. Mov. Sci.* 44, 32–41. doi: 10.1016/j.humov.2015.08.011

Amano, S., and Kondo, T. (1999). *NTT Database Series, Nihongo-no Goitokusei: Lexical Properties of Japanese (Vol. 1)*. Tokyo: Sanseido-shoten Co. Ltd.

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., et al. (2014). Video game training enhances cognitive control in older adults. *Nature* 501, 97–101. doi: 10.1038/nature12486

Bakic, J., Jepma, M., De Raedt, R., and Pourtois, G. (2014). Effects of positive mood on probabilistic learning: behavioral and electrophysiological correlates. *Biol. Psychol.* 103, 223–232. doi: 10.1016/j.biopsycho.2014.09.012

Borghini, G., Vecchiato, G., Toppi, J., Astolfi, L., Maglione, A., Isabella, R., et al. (2012). Assessment of mental fatigue during car driving by using high resolution EEG activity and neurophysiologic indices. *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 70, 6442–6445. doi: 10.1109/EMBC.2012.6347469

Chwilla, D. J., Virgillito, D., and Vissers, C. T. W. M. (2011). The relationship of language and emotion: N400 support for an embodied view of language comprehension. *J. Cogn. Neurosci.* 23, 2400–2414. doi: 10.1162/jocn.2010.21578

Deacon, D., Dynowska, A., Ritter, W., and Grose-Fifer, J. (2004). Repetition and semantic priming of nonwords: implications for theories of N400 and word recognition. *Psychophysiology* 41, 60–74. doi: 10.1111/1469-8986.00120

Federmeier, K. D., Kirson, D. A., Moreno, E. M., and Kutas, M. (2001). Effects of transient, mild mood states on semantic memory organization and use: an event-related potential investigation in humans. *Neurosci. Lett.* 305, 149–152. doi: 10.1016/s0304-3940(01)01843-2

Friedrich, M., and Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: processing known words in picture contexts. *J. Cogn. Neurosci.* 16, 1465–1477. doi: 10.1162/0898929042304705

Friedrich, M., and Friederici, A. D. (2010). Maturing brain mechanisms and developing behavioral language skills. *Brain Lang.* 114, 66–71. doi: 10.1016/j.bandl.2009.07.004

Gerth, S., Dolk, T., Klassert, A., Fliesser, M., Fischer, M. H., Nottbusch, G., et al. (2016a). Adapting to the surface: a comparison of handwriting measures when writing on a tablet computer and on paper. *Hum. Mov. Sci.* 48, 62–73. doi: 10.1016/j.humov.2016.04.006

Gerth, S., Klassert, A., Dolk, T., Fliesser, M., Fischer, M. H., Nottbusch, G., et al. (2016b). Is handwriting performance affected by the writing surface? Comparing preschoolers', second graders', and adults' writing performance on a tablet vs. paper. *Front. Psychol.* 7:1308. doi: 10.3389/fpsyg.2016.01308

Gevins, A., Smith, M. E., McEvoy, L., and Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cereb. Cortex* 7, 374–385. doi: 10.1093/cercor/7.4.374

Guilbert, J., Alamargot, D., and Morin, M. F. (2019). Handwriting on a tablet screen: role of visual and proprioceptive feedback in the control of movement by children and adults. *Hum. Mov. Sci.* 65, 30–41. doi: 10.1016/j.humov.2018.09.001

Hatano, A., Sekine, T., Herai, T., Ihara, N., Tanaka, Y., Murakami, S., et al. (2015). Effects of the use of paper notebooks and tablet devices on cognitive load in learning-An Electroencephalographic (EEG) study. *IEICE Technic. Rep.* 115, 39–44.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: implications for the role of the N400 in language processing. *Psychophysiology* 30, 47–61. doi: 10.1111/j.1469-8986.1993.tb03204.x

Kiefer, M., Schuler, S., Mayer, C., Trumpp, N. M., Hille, K., and Sachse, S. (2015). Handwriting or typewriting? The influence of pen-or keyboard-based writing training on reading and writing performance in preschool children. *Adv. Cogn. Psychol.* 11, 136–146. doi: 10.5709/acp-0178-7

Kutas, M., and Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470. doi: 10.1016/s1364-6613(00)01560-6

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Longcamp, M., Gilhodes, J., Anton, J., Roth, M., Nazarian, B., and Velay, J. (2008). Learning through hand-or typewriting influences visual recognition of new graphic shapes: behavioral and functional imaging evidence. *J. Cogn. Neurosci.* 20, 802–815. doi: 10.1162/jocn.2008.20504

Longcamp, M., Zerbato-Poudou, M. T., and Velay, J. L. (2005). The influence of writing practice on letter recognition in preschool children: a comparison between handwriting and typing. *Acta Psychol.* 119, 67–79. doi: 10.1016/j.actpsy.2004.10.019

Mangen, A., Anda, L. G., Oxborough, G. H., and Brønnick, K. (2015). Handwriting versus keyboard writing: effect on word recall. *J. Writ. Res.* 7, 227–247. doi: 10.17239/jowr-2015.07.02.1

Matsumoto, A., Iidaka, T., Haneda, K., Okada, T., and Sadato, N. (2005). Linking semantic priming effect in functional MRI and event-related potentials. *Neuroimage* 24, 624–634. doi: 10.1016/j.neuroimage.2004.09.008

McLaughlin, J., Osterhout, L., and Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nat. Neurosci.* 7, 703–704. doi: 10.1038/nn1264

Mueller, P. A., and Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: advantages of longhand over laptop note taking. *Psychol. Sci.* 25, 1159–1168. doi: 10.1177/0956797614524581

Nadler, R. T., Rabi, R., and Minda, J. P. (2010). Better mood and better performance. Learning rule-described categories is enhanced by positive mood. *Psychol. Sci.* 21, 1770–1776. doi: 10.1177/0956797610387441

Ojima, S., Nakamura, N., Matsuba-Kurita, H., Hoshino, T., and Hagiwara, H. (2011). Neural correlates of foreign-language learning in childhood: a 3-year longitudinal ERP study. *J. Cogn. Neurosci.* 23, 183–199. doi: 10.1162/jocn.2010.21425

Ojima, S., Nakata, H., and Kakigi, R. (2005). An ERP study of second language learning after childhood: effects of proficiency. *J. Cogn. Neurosci.* 17, 1212–1228. doi: 10.1162/0898929055002436

Reid, V. M., Hoehl, S., Grigutsch, M., Groendahl, A., Parise, E., and Striano, T. (2009). The neural correlates of infant and adult goal prediction: evidence for semantic processing systems. *Dev. Psychol.* 45, 620–629. doi: 10.1037/a0015209

Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. *Phychophysiology* 22, 642–647. doi: 10.1111/j.1469-8986.1985.tb01661.x

Sakuma, N., Ijuin, M., Fushimi, T., Tatsumi, I., Tanaka, M., Amano, S., et al. (2005). *Nihongo-No Goitokusei: Lexical Properties of Japanese (Vol.8).* Tokyo: Sanseido-shoten Co. Ltd.

van der Meer, A. L. H., and van der Weel, F. R. (2017). Only three fingers write, but the whole brain works: a high-density EEG study showing advantages of drawing over typing for learning. *Front. Psychol.* 8:706. doi: 10.3389/fpsyg.2017.00706

van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., and Mcisaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *J. Cogn. Neurosci.* 3, 131–150. doi: 10.1162/jocn.1991.3.2.131

Vinci-booher, S., James, T. W., and Karin, H. (2016). Visual-motor functional connectivity in preschool children emerges after handwriting experience. *Trends Neurosci. Educ.* 5, 107–120. doi: 10.1016/j.tine.2016.07.006

Vissers, C. T. W. M., Chwilla, U. G., Egger, J. I. M., and Chwilla, D. J. (2013). The interplay between mood and language comprehension: evidence from P600 to semantic reversal anomalies. *Neuropsychologia* 51, 1027–1039. doi: 10.1016/j.neuropsychologia.2013.02.007

Vissers, C. T. W. M., Virgillito, D., Fitzgerald, D. A., Speckens, A. E. M., Tendolkar, I., van Oostrom, I., et al. (2010). The influence of mood on the processing of syntactic anomalies: evidence from P600. *Neuropsychologia* 48, 3521–3531. doi: 10.1016/j.neuropsychologia.2010.08.001

Wollscheid, S., Sjaastad, J., and Tømte, C. (2016). The impact of digital devices vs. Pen(cil) and paper on primary school students' writing skills—a research review. *Comput. Educ.* 95, 19–35. doi: 10.1016/j.compedu.2015.12.001

Conflict of Interest Statement: AK and KI are employees of Wacom Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 20 December 2019
doi: 10.3389/fnhum.2019.00449

Occupational Patterns of Structural Brain Health: Independent Contributions Beyond Education, Gender, Intelligence, and Age

Christian Habeck, Teal S. Eich, Yian Gu and Yaakov Stern*

Cognitive Neuroscience Division, Department of Neurology, Columbia University, New York, NY, United States

Occupational activity represents a large percentage of people's daily activity and thus likely is as impactful for people's general and cognitive health as other lifestyle components such as leisure activity, sleep, diet, and exercise. Different occupations, however, require different skills, abilities, activities, credentials, work styles, etc., constituting a rich multidimensional formative exposure with likely consequences for brain development over the lifespan. In the current study, we were interested in how different occupations with their different attributes relate to five variables: structural brain health, duration of early-life education, gender, IQ, and age, although the main focus was the relationship to brain health. To this end, we used the Occupation Information Network (O*NET), which provides quantification of occupations along 246 items. Occupational patterns with different loadings for these 246 items were derived from 277 community-dwelling adults, ranging in age from 40 to 80, based upon the five subject measures. We found significant patterns underlying four of our variables of interest, with gender and education predictably showing the most numerous and strongest associations, while brain health and intelligence showed weaker associations, and age did not manifest any associations. For the occupational pattern associated with brain health, we found mainly positive associations on items pertaining to rigorous problem-solving, leadership, responsibility, and information processing. We emphasize that the findings are correlational and cannot establish causation. Future extensions of this work will assess the influence of occupation on future cognitive brain status and cognitive performance.

Keywords: cortical thickness, occupational data, community cohort, education, age, gender

OPEN ACCESS

Edited by:
Michael Valenzuela,
The University of Sydney, Australia

Reviewed by:
Roger Staff,
NHS Grampian, United Kingdom
Chao Suo,
Monash University, Australia

****Correspondence:***
Christian Habeck
ch629@columbia.edu;
ch629@cumc.columbia.edu

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *30 August 2019*
Accepted: *05 December 2019*
Published: *20 December 2019*

Citation:
Habeck C, Eich TS, Gu Y and Stern Y (2019) Occupational Patterns of Structural Brain Health: Independent Contributions Beyond Education, Gender, Intelligence, and Age. Front. Hum. Neurosci. 13:449. doi: 10.3389/fnhum.2019.00449

INTRODUCTION

Occupational attainment and fulfillment has been linked to successful cognitive performance (Garibotto et al., 2008, 2012; Bickel and Kurz, 2009; Foubert-Samier et al., 2012; Pool et al., 2016; Chan et al., 2018; Dodich et al., 2018), psychiatric aging, and general well-being (Dragano et al., 2011; Platts et al., 2013; Wahrendorf et al., 2013) in older adults, independent of socio-economic and educational status.

The relation between occupational attainment/satisfaction and markers of brain-structural health, however, has been probed and observed far less frequently. Some studies have addressed relationships between occupation and brain metabolism (Spreng et al., 2010; Spreng et al., 2011) and white-matter tract integrity (Kaup et al., 2018). In the few studies that have examined cortical thickness and volume, occupation has been shown to be negatively associated with occupational stress (Blix et al., 2013; Savic, 2015; Savic et al., 2018). Negative associations after controlling for clinical disease severity in neurodegenerative disease suggests that occupational attainment is a form of cognitive reserve (Boots et al., 2015).

In the current study, we were interested in the relationship between occupation and structural brain health, with a particular interest in the extent of this relationship beyond possible demographic variables that are collinear confounders of general health status, including age, education, IQ, and gender.

We studied and used occupational attainment as quantified by the extensive characterization in the Occupational Information Network (O*NET[1]), an online resource maintained by the US Department of Labor. Every job, for example Physics Teacher, Postsecondary or Marketing Manager, is assigned a Standard Occupational Classification (SOC) numeric code, 25-1054.00 and 11-2021.00 respectively, and is quantified in terms of a multitude of indicator variables or dimensions. Here, we followed previous research conventions (Peterson and American Psychological Association, 1999; Gadermann et al., 2014) and retained 246 worker-centric variables with data-ratings. These dimensions are drawn from different domains, including 'work values' (6 items), 'interests' (6 items), 'knowledge' (33 items), 'abilities' (52 items), 'work activities' (41 items), 'work styles' (16 items), 'skills' (35 items), and 'work context' (57 items). To further illustrate this taxonomy, we give a few examples of items in these different categories. 'Work values,' 'interests,' and 'knowledge' are more general and abstract, and thus the labels need more explicit consultation of the online data base. For instance, 'work values' concerns items such as 'achievement' which specifies an orientation awards results and accomplishments, whereas 'support' captures occupations that involve institutionalized support structures (management, HR, etc.). A watch repairer (code 49-9064.00), for instance, would score high on 'achievement,' but low on 'support.'

Domains that are more concrete are 'work styles,' 'work context,' 'skills,' 'work activities,' and 'abilities.' The labels for the these items are usually self-explanatory, such as 'Persistence' (work styles), 'Contact with others' (work context),' 'Science' (skills), 'Interacting with computers' (work activities), or 'Memorization' (abilities). For all results in this paper, every item label will also be supplemented by the appropriate domain label. If the label is not self-explanatory, the exact definition can be looked up in O*NET.

It is noteworthy that the quantification of occupations along the 246 dimensions necessarily induces positive or negative correlations between items. The reasons are twofold: (1) some of the items are intrinsically similar or oppositional. For instance, the complementary work-context item 'Time spent standing' can only correlate negatively with 'Time spent sitting.' (2) More interestingly, some items are not intrinsically similar or oppositional, but they become so because of the empirical nature of most occupations in our sample. The skill- and work-activities items 'Critical thinking' and 'Handling and moving objects' are not *a priori* oppositional, and there might be specialized occupations that require both. However, in our sample – and probably the majority of population-based research– they are negatively correlated ($R = -0.51$, $p < 0.0001$). Further, occupational data will most likely be rank-deficient, i.e., the numbers of observations (= participants in sample) might be larger than the number of different occupations. This is the case for our data array, where 277 participants constitute only 152 different occupations.

Apart from inherent correlations between the occupational items, occupational attainment, intelligence, education and brain structural health also usually show mutual associations, and this was no different in our data. Thus, it is difficult isolate a relation between brain health and occupation free from these confounders in cross-sectional associations. At the same time, randomized interventions with occupation are either impossible, or at least only possible in very narrow contexts, and so associational studies have to resort to techniques that try to adjust for the confounders *post hoc*.

In the current study, we investigated the association between a measure of structural brain health and occupational attainment in 246 indicator variables in a community-based cohort of 277 participants, aged 40 to 80. Gender, age, education, and IQ were simultaneously entered with structural brain health as covariates in a general linear model to identify associated items in the occupational data.

MATERIALS AND METHODS

Subject Sample, Acquired Data, and Pre-processing

Participants who lived within a radius of 10 miles of the Columbia University Medical Center were recruited to the study via random market mailing, and were screened for magnetic resonance imaging (MRI) contraindications and hearing or visual impairment that would impede testing. Older adult participants were additionally screened for dementia and mild cognitive impairment prior to participating in the study, and participants who met criteria for either were excluded. Apart from these cognitive exclusion criteria, health-related exclusion criteria included myocardial infarction, congestive heart failure or any other heart disease, brain disorder such as stroke, tumor, infection, epilepsy, multiple sclerosis, degenerative diseases, head injury (loss of consciousness > 5 min), intellectual disability, seizure, Parkinson's disease, Huntington's disease, normal pressure hydrocephalus, essential/familial tremor, Down Syndrome, HIV Infection or AIDS diagnosis, learning disability/dyslexia, ADHD or ADD, uncontrolled hypertension, uncontrolled diabetes mellitus, uncontrolled thyroid or other endocrine disease, uncorrectable vision, color blindness,

[1]https://www.onetonline.org/

uncorrectable hearing and implant, pregnancy, lactating, any medication targeting central nervous system, cancer within last 5 years, renal insufficiency, untreated neurosyphillis, any alcohol and drug abuse within last 12 month, recent non-skin neoplastic disease or melanoma, active hepatic disease, insulin dependent diabetes, any history of psychosis or ECT, recent (past 5 years) major depressive, bipolar, or anxiety disorder, objective cognitive impairment (dementia rating scale of < 130), and subjective functional impairment (BFAS > 1).

All procedures undertaken for this study were approved by the Columbia Institutional Review Board. **Table 1** provides sample information. IQ was assessed with the National Adult Reading Test (Nelson, 1982; Blair and Spreen, 1989).

Occupational Data Acquisition

Comprehensive EXCEL spreadsheets for all 8 domain labels for 969 occupations were downloaded from O*NET in August 2016, then processed and collated following prior established convention (Gadermann et al., 2014). Participants were asked to provide the occupation of the longest duration during their lifetime. A Research Assistant matched the occupation to the O*NET SOC code, and the 246 indicator variables for each code were obtained from the collated spreadsheet.

Structural Brain Data Acquisition (T1, DTI, and FLAIR) and Processing

Magnetic resonance imaging images were acquired in a 3.0T Philips Achieva Magnet using a standard quadrature head coil. A T1-weighted scout image was acquired to determine subject position. One hundred sixty-five contiguous 1 mm coronal T1-weighted images of the whole brain were acquired for each subject with an MPRAGE sequence using the following parameters: TR 6.5 ms, TE 3 ms; flip angle 8°, acquisition matrix 256 × 256 and 240 mm field of view. The DTI images were acquired in 55 directions using these parameters: $b = 800$ s/mm^2, TE = 69 ms, TR = 11032 ms, Flip Angle = 90°, in-plane resolution 112 × 112 voxels, acquisition time 12 min 56 s, slice thickness = 2 mm (no gap), 75 slices. Lastly, a FLAIR scan was acquired with the following parameters: 11,000 ms TR, 2800 ms TE, 256 × 189 voxels in-plane resolution, 23.0 × 17.96 cm field of view (FOV), and 30 slices with slice-thickness/gap of 4/0.5 mm. This sequence was used to quantify the WMHs volumes. A neuroradiologist reviewed each scan individually to exclude any relevant findings. In the case of a clinical positive finding, the subject's primary care physician was informed.

TABLE 1 | Number, age, years of education, and IQ of the participant sample.

	Participant sample
Age, mean ± STD, range	61.76 ± 9.34, 40–78
Total number, women, men	277, 142 W, 135 M
Self-identified race	67 African American, 6 Asian, 189 Caucasian, 2 Pacific Islander, 2 Mixed Race, 11 Other
Education in years, mean ± STD, range	16.31 ± 2.39, 12–22
IQ, mean ± STD, range	118.52 ± 8.99, 93.60–130.88

IQ was assessed with the National Adult Reading Test.

Each subject's structural T1 scans were reconstructed using FreeSurfer v5.1[2]. The accuracy of FreeSurfer's subcortical segmentation and cortical parcelation (Fischl et al., 2002, 2004) has been reported to be comparable to manual labeling. Each subject's white and gray matter boundaries, as well as gray matter and cerebral spinal fluid boundaries, were visually inspected slice by slice, and manual control points were added in the case of any visible discrepancy. Reconstruction was repeated until we reached satisfactory results within every subject. The subcortical structure borders were plotted by *freeview* visualization tools and compared against the actual brain regions. In case of discrepancy, they were corrected manually. Finally, we obtained cortical thickness for 68 regions of nterest (ROIs), and also read out the main global-thickness value provided by FreeSurfer.

DTI data were processed with TRACULA (Tracts Constrained by Underlying Anatomy) distributed as part of the FreeSurfer v. 5.2 library (Yendiki et al., 2011) which produces 18 major White-Matter tracts. The software performs informed automatic tractography by incorporating anatomical information from a training data set, provided by the software, with the anatomical segmentation of the T1 image of the current data set, thus increasing the accuracy of the WM tract placement for each participant. Standard DTI processing steps using the FMRIB's Diffusion Toolbox (FMRIB's Software Library v. 4.1.5) including eddy current correction, tensor estimation, and bedpostx were performed prior to tractography by the TRACULA software (Yendiki et al., 2011). For each participant, the means of fractional anisotropy (FA) for each of the 18 tracts, were entered into subsequent analyses. FA ranges from 0 to 1 with higher number representing more intact WM integrity.

White-Matter-Hyperintensities (WMH) were obtained through segmentation by the Lesion Segmentation Tool algorithm (LST) (Schmidt et al., 2012) as implemented in the LST toolbox version 2.0.15 (June 2017) for Statistical Parametric Mapping (SPM)[3]. The algorithm first segments the T1 images into the three main tissue classes – cerebral brain fluid, gray matter and white matter. Then, this information is combined with the co-registered FLAIR intensities in order to calculate lesion belief maps. By thresholding these maps with a pre-chosen initial threshold, an initial binary lesion map is obtained which is subsequently grown along voxels that appear hyper intense in the FLAIR image. The result is a lesion probability map. Every FLAIR sequence that had a total WMH volume above 1000 mm^3 was manually inspected to ensure that there were no visible discrepancies. We counted the number of hyper-intense voxels, N, that were classified as hyper intense and transformed as log-WMH = log(N + 1).

[2]http://surfer.nmr.mgh.harvard.edu/

[3]www.statistical-modelling.de/lst.html

Data Analysis: Multimodal Brain Health Computation

Structural brain health was computed as the multimodal average of global cortical thickness (the total value provided by FreeSurfer, *not* the average of the 68 ROIs), mean tract integrity and the sign-reversed log-WMH measure. Since the three constituents are incommensurate, they were first z-scored and then averaged according to

$$\text{brain health} = [\text{z(global thickness)} + \text{z(mean tract integrity)} - \text{z(log-WMH)}]/3.$$

We note that our operationalization of this measure is just an obvious starting point in terms of simplicity, but other formulations are conceivable too. Other modalities with differential contributions might be added, optimized for considerations of construct validity beyond this study. Supplementary regression models were run where the brain-health variable was substituted by individual cortical thicknesses. Results can be found in **Supplementary Table S2**.

Data Analysis: Mass-Univariate Analysis

We first performed mass-univariate analysis by simultaneously entering all covariates (brain health, education, IQ, gender, age, and race) and performing a linear regression according to

$$\text{occ(i)} = \left[\text{brain-health education IQ gender age race } 1\right]\beta + \varepsilon \quad i = 1 \ldots 246.$$

with a False-Discovery Rate (FDR) of Q < 0.05 (Hochberg and Benjamini, 1990). (**1** denotes the intercept term.) Race was coded as a categorical index array with values of 0 or 1, had 2 columns and 277 rows. Column 1 indicated the status of 'African American' ($N = 67$), and column 2 combined the labels 'Mixed Race,' 'Asian,' 'Pacific Islander,' and 'Other' ($N = 21$), and thus could be labeled as 'Neither African American nor Caucasian.'

RESULTS

We first ran our mass-univariate linear regression with the full covariate set including the racial index array. However, we did not identify any associations between occupation items and race at Q < 0.05, and decided to drop the racial index array from our analyses to increase statistical power. To arrival at our final results, we re-ran the regression models with the reduced set of five covariates: (1) brain health, (2) education, (3) gender, (4) NART-IQ, and (5) age.

Collinearity of Covariates

To convey an impression of the collinearity of the covariates we report all bivariate correlations at an uncorrected p-value of $p < 0.05$. Age displays an expected strong negative correlation with total brain health ($R = -0.54$, $p < 0.0001$) and positive associations with education ($R = 0.12$, $p = 0.04$), NART-IQ ($R = 0.18$, $p = 0.0025$). Lastly, as expected, NART-IQ and education are highly correlated at $R = 0.54$, $p < 0.0001$.

Univariate Analysis With FDR Correction

We found significant associations at $Q < 0.05$ for all covariates except age. We first turn our attention to the main objective of this study: brain health. We list the 10 strongest associations in **Table 2**, but give a full listing of the occupational profiles for all covariates in **Supplementary Table S1**.

The items associated with brain health (above and beyond the other covariates) contain a mixture of all domain labels apart from 'Interests.' Inspection of all positively correlated items shows work activities, styles and context show items that involve processing of information, numerical reasoning and decision making with the help of computers, facing responsibility and having to show leadership with severe consequence of errors. Numerical and critical-thinking skills and abilities were strongly associated with better brain health too, as were work styles that emphasizes persistence, initiative and leadership. The knowledge item 'Foreign Language' showed the only negative association.

For the other covariates, education by far showed the most numerous and significant associations with 180 items (see **Supplementary Table S1**). We display an abbreviated listing in **Table 3**, giving the first 10 items in both directions of association.

Table 3 and the full listing in **Supplementary Table S1** show that items pertaining to work-context, -activities, skills and abilities associated with manual labor show a negative association with education, while items associated with white-collar knowledge work are associated positively with education.

Gender shows similarly strong effects, probably expressing stereotypical gender roles with occupation choice that – over time- might reduce. Women choose occupations that show more traditionally female attributes with little

TABLE 2 | Abbreviated listing of up to 10 associations for occupational items and brain health at Q < 0.05.

Brain health – 39 items in total			
Item	**Domain**	***T***	***p***
Positive associations			
AnalyticalThinking	WorkStyles	4.0766	6.01E−05
InformationOrdering	Abilities	3.9901	8.50E−05
AchievementEffort	WorkStyles	3.9348	0.00010586
IdentifyingObjectsActionsandEvents	WorkActivities	3.8641	0.00013957
Support	WorkValues	3.7927	0.00018378
SystemsAnalysis	Skills	3.6278	0.0003415
Mathematics	Knowledge	3.5921	0.00038941
CriticalThinking	Skills	3.5737	0.00041658
MonitorProcessesMaterialsor Surroundings	WorkActivities	3.4841	0.00057573
ComplexProblemSolving	Skills	3.4362	0.00068262
Negative associations			
ForeignLanguage	Knowledge	−2.7505	0.0063503

There was only a single negative association, but 38 positive associations in total. The p-values listed are uncorrected.

TABLE 3 | Abbreviated list of items displaying significant correlations with education at Q < 0.05.

Education – 180 items in total			
Item	**Domain**	***T***	***p***
Positive associations			
WrittenComprehension	Abilities	7.5256	7.78E−13
ActiveLearning	Skills	7.2708	3.84E−12
OralExpression	Abilities	7.2525	4.30E−12
WrittenExpression	Abilities	7.1636	7.43E−12
ReadingComprehension	Skills	7.1374	8.73E−12
Writing	Skills	7.0934	1.14E−11
Speaking	Skills	6.9272	3.12E−11
JudgmentandDecisionMaking	Skills	6.6831	1.33E−10
OralComprehension	Abilities	6.618	1.94E−10
DeductiveReasoning	Abilities	6.4681	4.61E−10
Negative associations			
SpendTimeKneelingCrouching StoopingorCrawling	WorkContext	−5.9534	8.11E−09
StaticStrength	Abilities	−5.7125	2.93E−08
MultilimbCoordination	Abilities	−5.5934	5.44E−08
SpendTimeBendingorTwistingtheBody	WorkContext	−5.5346	7.36E−08
CrampedWorkSpaceAwkwardPositions	WorkContext	−5.4708	1.02E−07
HandlingandMovingObjects	WorkActivities	−5.4265	1.28E−07
ManualDexterity	Abilities	−5.3087	2.30E−07
SpeedofLimbMovement	Abilities	−5.2865	2.57E−07
ExtentFlexibility	Abilities	−5.2267	3.45E−07
GrossBodyCoordination	Abilities	−5.0671	7.49E−07

The p-values listed are uncorrected.

TABLE 4 | Abbreviated list of items displaying significant correlations with gender at Q < 0.05.

Gender – 57 items in total			
Item	**Domain**	***T***	***p***
Positive associations (i.e., associations with being female)			
Artistic	Interests	3.5895	0.00039319
Independence	WorkStyles	3.5324	0.00048385
SocialOrientation	WorkStyles	3.5204	0.00050533
Innovation	WorkStyles	3.4141	0.00073776
FineArts	Knowledge	3.3917	0.00079826
SociologyandAnthropology	Knowledge	3.3478	0.00092996
CommunicationsandMedia	Knowledge	3.3209	0.0010205
PhilosophyandTheology	Knowledge	3.3119	0.0010524
Clerical	Knowledge	3.3113	0.0010547
Dependability	WorkStyles	3.2894	0.0011368
Negative associations (i.e., associations with being male)			
SoundLocalization	Abilities	−4.3912	1.62E−05
SpatialOrientation	Abilities	−4.2588	2.84E−05
InanOpenVehicleorEquipment	WorkContext	−4.2431	3.03E−05
NightVision	Abilities	−4.1894	3.79E−05
SpendTimeClimbingLaddersScaffoldsor Poles	WorkContext	−4.101	5.44E−05
Mechanical	Knowledge	−4.0904	5.68E−05
PeripheralVision	Abilities	−4.0362	7.07E−05
OperatingVehiclesMechanizedDevicesor Equipment	WorkActivities	−3.9368	0.00010506
GlareSensitivity	Abilities	−3.8362	0.00015547
Realistic	Interests	−3.7558	0.00021151

Because gender is coded as Female = 2, Male = 1, positive correlations pertain to items associated with female gender, negative correlations pertain to items associated with male gender. The p-values listed are uncorrected.

constraint by work context, whereas men preferentially have occupations that involve technical expertise, sensory-perception demands and manual labor. We give the abbreviated listing in **Table 4**.

Lastly, we list the items associated with crystallized intelligence, i.e., NART-IQ, in **Table 5** in full. There were only eight items in total.

After deriving the occupational profiles of all covariates, we decided to inspect the similarity between the brain-health occupational profile and all remaining profiles with simple bivariate scatter plots (see **Figure 1**). This second-order correlation can at least visualize the similarity of the occupation-covariate relationships in relative terms. Interestingly, the brain-health profile shows the greatest similarity to the profiles of education and age (although no individual occupational item showed an association with age at $Q < 0.05$). This similarity is present although the covariates brain health and education showed no relationship, while brain health and age showed a strong *negative* relationship. In our sample at least, older participants chose occupations that are also associated with better brain health and higher education.

The gender-associated occupational profile showed no relationship to the brain-health profile, while the NART-IQ-related profile showed a weak *negative* relationship. At the level of covariates, brain health was unrelated to either gender or NART-IQ.

DISCUSSION

The main purpose of this study was to clarify the fine-grained relationship between structural brain health and occupation, adjusted for education, gender, age, and IQ. We emphasize again that the results are correlational, and that no inference regarding causal directionality can be made.

The occupational profiles of education, sex, and NART-IQ were somewhat in line with common-sense expectations which would attribute manual-labor and sensory-perception skills and abilities predominantly to male or lower educated participants, while items associated with social orientation, higher knowledge, fine arts and communication were differentially and independently associated with being female and more educated.

For our main association of interest, we mainly found positive associations between items pertaining to processing of information, numerical reasoning, problem-solving and decision making with the help of computers, facing responsibility and having to show leadership with severe consequence of errors. Further correlations were shown with numerical and critical-thinking skills and abilities, and with work styles that emphasize persistence, responsibility, initiative and leadership. Brain health in our operationalization did not show any confounding correlations with education and gender; further, when we ran supplementary analyses leaving out gender and education,

TABLE 5 | Full listing of items associated with NART-IQ at Q < 0.05.

NART-IQ – 8 items in total			
Item	**Domain**	***T***	***p***
Positive associations			
Artistic	Interests	5.1803	4.33E-07
FineArts	Knowledge	4.5252	9.04E-06
Innovation	WorkStyles	4.1979	3.66E-05
ThinkingCreatively	WorkActivities	3.6928	0.00026824
Originality	Abilities	3.5274	0.0004928
Negative associations			
Telephone	WorkContext	−3.6853	0.00027582
Integrity	WorkStyles	−3.4688	0.00060787

The p-values listed are uncorrected.

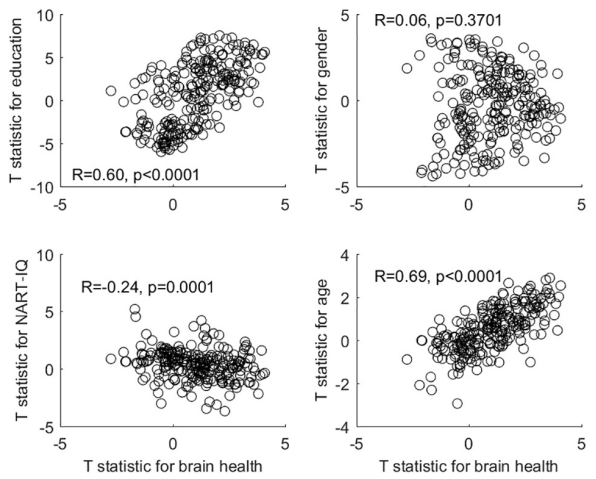


FIGURE 1 | Bivariate plots and correlations between the occupational profile (= T-statistic) of brain health and the occupational profiles of all other covariates.

the items that were recovered with significant associations were very similar and – in fact-fewer in number. Thus, our results indicate robust relationships between occupation and brain health that are not fully mediated by education, intelligence, or gender.

Furthermore, higher-order correlation of the occupational profiles (= T-statistic) across all 246 items revealed that the brain-health profile was similar (in the sense of being significantly correlated) to the education profile, despite both covariates sharing no significant relationship. There was a likewise similarity between the brain-health occupational profile and the age-related occupational profile, even though brain health and age are strongly *negatively* associated (and no individual item in the age profile reached statistical significance at $Q < 0.05$).

Several caveats must be mentioned in our study design: (1) important information about parental socio-economics and upbringing were missing, although these factors are certain to influence brain development (Noble et al., 2015) beyond the duration of early-life education. To arrive at a relationship between occupation and brain health, this confounder would have to be taken into account. (2) While education and occupation are *not* contemporaneous with, and predate, the brain-health assessment, it is tempting to speculate about causal relationships. It could be that some occupational demands serve as cognitive training regimens that result in better brain health, while some job aspects (particularly environmental exposures) could be detrimental to brain health. However, even for cross-sectional correlations, long-lasting influences of other factors (such as parenting style and early-life socio-economics) would have to be taken into account. To reduce the possibility of reverse causation, i.e., brain health at an early age leading to particular educational and occupational choices, brain health at a young age ideally should also be considered. (3) We only recorded the occupation with the longest tenure in our participants' lives, and no more detailed information about occupation sequences were queried.

We close our report with some suggestions for future extensions, sparked by the study limitations: a more complete record of occupational history and parental socio-economics is indispensable for a refinement of the relationship between occupation and brain health. Further, while interventional studies for occupation are hard to conceive, prospective cohort studies could record more complete and dynamic occupational information and establish relationships to future brain structural measures, thus getting closer to true a causal account. As mentioned in the introductory remarks, the large amount of time that work represents in the daily routine for most people suggests that occupational choices and demands would be reflected in the brain, similar to other lifestyle features such as exercise, diet, sleep, and leisure activities. To clarify the role of occupation for better brain maintenance and cognitive reserve will be an exciting endeavor in brain research for the foreseeable future.

Our study also hopes to introduce the O*NET database to a broader audience and convey some of the benefits of the fine-grained quantitative assessment of occupation. We only performed simple univariate analyses, which is a natural starting point. O*NET enables more sophisticated frameworks of course, and gives the opportunity of operationalizing similarity and 'distance' between occupations, with multivariate decompositions of occupational profiles that capture dimensions other than education, gender, and intelligence. Occupational data might provide a fertile ground for identifying factors with predictive utility for prognosis and diagnosis of cognitive dysfunction in addition to structural brain markers and age.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher. A MATLAB data archive is available with annotation from the corresponding author on request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Columbia University IRB. The

patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CH conceived of the study, analyzed the data, and wrote the draft. TE performed supplementary data analysis and formatting of the occupational data. YG read the manuscript, performed supplementary data analysis, and gave feedback. YS discussed the study extensively, read the draft, and provided commentary.

FUNDING

Support from grants NIH/NIA RF1AG038465 and R01AG026158 is gratefully acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2019.00449/full#supplementary-material

REFERENCES

Bickel, H., and Kurz, A. (2009). Education, occupation, and dementia: the Bavarian school sisters study. *Dement. Geriatr. Cogn. Disord.* 27, 548–556. doi: 10.1159/000227781

Blix, E., Perski, A., Berglund, H., and Savic, I. (2013). Long-term occupational stress is associated with regional reductions in brain tissue volumes. *PLoS One* 8:e64065. doi: 10.1371/journal.pone.0064065

Boots, E. A., Schultz, S. A., Almeida, R. P., Oh, J. M., Koscik, R. L., Dowling, M. N., et al. (2015). Occupational complexity and cognitive reserve in a middle-aged cohort at risk for Alzheimer's Disease. *Arch. Clin. Neuropsychol.* 30, 634–642. doi: 10.1093/arclin/acv041

Chan, D., Shafto, M., Kievit, R., Matthews, F., Spink, M., Valenzuela, M., et al. (2018). Lifestyle activities in mid-life contribute to cognitive reserve in late-life, independent of education, occupation, and late-life activities. *Neurobiol. Aging* 70, 180–183. doi: 10.1016/j.neurobiolaging.2018.06.012

Dodich, A., Carli, G., Cerami, C., Iannaccone, S., Magnani, G., and Perani, D. (2018). Social and cognitive control skills in long-life occupation activities modulate the brain reserve in the behavioural variant of frontotemporal dementia. *Cortex* 99, 311–318. doi: 10.1016/j.cortex.2017.12.006

Dragano, N., Siegrist, J., and Wahrendorf, M. (2011). Welfare regimes, labour policies and unhealthy psychosocial working conditions: a comparative study with 9917 older employees from 12 European countries. *J. Epidemiol. Community Health* 65, 793–799. doi: 10.1136/jech.2009.098541

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355.

Fischl, B., Salat, D. H., van der Kouwe, A. J., Makris, N., Segonne, F., Quinn, B. T., et al. (2004). Sequence-independent segmentation of magnetic resonance images. *Neuroimage* 23(Suppl. 1), S69–S84.

Foubert-Samier, A., Catheline, G., Amieva, H., Dilharreguy, B., Helmer, C., Allard, M., et al. (2012). Education, occupation, leisure activities, and brain reserve: a population-based study. *Neurobiol Aging* 33, 423.e15–423.e25. doi: 10.1016/j.neurobiolaging.2010.09.023

Gadermann, A. M., Heeringa, S. G., Stein, M. B., Ursano, R. J., Colpe, L. J., Fullerton, C. S., et al. (2014). Classifying U.S. Army Military Occupational Specialties using the Occupational Information Network. *Mil. Med.* 179, 752–761. doi: 10.7205/MILMED-D-13-00446

Garibotto, V., Borroni, B., Kalbe, E., Herholz, K., Salmon, E., Holtoff, V., et al. (2008). Education and occupation as proxies for reserve in aMCI converters and AD: FDG-PET evidence. *Neurology* 71, 1342–1349. doi: 10.1212/01.wnl.0000327670.62378.c0

Garibotto, V., Borroni, B., Sorbi, S., Cappa, S. F., Padovani, A., and Perani, D. (2012). Education and occupation provide reserve in both ApoE epsilon4 carrier and noncarrier patients with probable Alzheimer's disease. *Neurol. Sci.* 33, 1037–1042. doi: 10.1007/s10072-011-0889-5

Hochberg, Y., and Benjamini, Y. (1990). More powerful procedures for multiple significance testing. *Stat. Med.* 9, 811–818. doi: 10.1002/sim.4780090710

Kaup, A. R., Xia, F., Launer, L. J., Sidney, S., Nasrallah, I., Erus, G., et al. (2018). Occupational cognitive complexity in earlier adulthood is associated with brain structure and cognitive health in midlife: the CARDIA study. *Neuropsychology* 32, 895–905. doi: 10.1037/neu0000474

Blair, J. R., and Spreen, O. (1989). Predicting premorbid IQ: a revision of the national adult reading test. *Clin. Neuropsychol.* 3, 129–136. doi: 10.1371/journal.pone.0205754

Nelson, H. E. (1982). *National Adult Reading Test (NART): Test Manual*. Windsor: NFER Nelson.

Noble, K. G., Houston, S. M., Brito, N. H., Bartsch, H., Kan, E., Kuperman, J. M., et al. (2015). Family income, parental education and brain structure in children and adolescents. *Nat. Neurosci.* 18, 773–778. doi: 10.1038/nn.3983

Peterson, N. G., and American Psychological Association, (1999). *An Occupational Information System for the 21st Century the Development of O*NET*. Washington, D. C: American Psychological Association.

Platts, L. G., Netuveli, G., Webb, E., Zins, M., Goldberg, M., Blane, D., et al. (2013). Physical occupational exposures during working life and quality of life after labour market exit: results from the GAZEL study. *Aging Ment. Health* 17, 697–706. doi: 10.1080/13607863.2013.781120

Pool, L. R., Weuve, J., Wilson, R. S., Bultmann, U., Evans, D. A., and Mendes de Leon, C. F. (2016). Occupational cognitive requirements and late-life cognitive aging. *Neurology* 86, 1386–1392. doi: 10.1212/WNL.0000000000002569

Savic, I. (2015). Structural changes of the brain in relation to occupational stress. *Cereb. Cortex* 25, 1554–1564. doi: 10.1093/cercor/bht348

Savic, I., Perski, A., and Osika, W. (2018). MRI Shows that exhaustion syndrome due to chronic occupational stress is associated with partially reversible cerebral changes. *Cereb. Cortex* 28, 894–906. doi: 10.1093/cercor/bhw413

Schmidt, P., Gaser, C., Arsic, M., Buck, D., Forschler, A., Berthele, A., et al. (2012). An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. *Neuroimage* 59, 3774–3783. doi: 10.1016/j.neuroimage.2011.11.032

Spreng, R. N., Drzezga, A., Diehl-Schmid, J., Kurz, A., Levine, B., and Perneczky, R. (2011). Relationship between occupation attributes and brain metabolism in frontotemporal dementia. *Neuropsychologia* 49, 3699–3703. doi: 10.1016/j.neuropsychologia.2011.09.025

Spreng, R. N., Rosen, H. J., Strother, S., Chow, T. W., Diehl-Schmid, J., Freedman, M., et al. (2010). Occupation attributes relate to location of atrophy in frontotemporal lobar degeneration. *Neuropsychologia* 48, 3634–3641. doi: 10.1016/j.neuropsychologia.2010.08.020

Wahrendorf, M., Blane, D., Bartley, M., Dragano, N., and Siegrist, J. (2013). Working conditions in mid-life and mental health in older ages. *Adv. Life Course Res.* 18, 16–25. doi: 10.1016/j.alcr.2012.10.004

Yendiki, A., Panneck, P., Srinivasan, P., Stevens, A., Zollei, L., Augustinack, J., et al. (2011). Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. *Front. Neuroinform.* 5:23. doi: 10.3389/fninf.2011.00023

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

The handling Editor declared a past co-authorship with one of the authors YS.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 10 January 2020
doi: 10.3389/fnhum.2019.00457

Revealing Relationships Among Cognitive Functions Using Functional Connectivity and a Large-Scale Meta-Analysis Database

Hiroki Kurashige[1,2], Jun Kaneko[1], Yuichi Yamashita[2], Rieko Osu[3], Yohei Otaka[4,5], Takashi Hanakawa[6], Manabu Honda[2] and Hideaki Kawabata[7]*

[1] Institute of Innovative Science and Technology, Tokai University, Tokyo, Japan, [2] National Institute of Neuroscience, National Center of Neurology and Psychiatry, Tokyo, Japan, [3] Faculty of Human Sciences, Waseda University, Tokyo, Japan, [4] Department of Rehabilitation Medicine I, School of Medicine, Fujita Health University, Aichi, Japan, [5] Department of Rehabilitation Medicine, Tokyo Bay Rehabilitation Hospital, Chiba, Japan, [6] Integrative Brain Imaging Center, National Center of Neurology and Psychiatry, Tokyo, Japan, [7] Department of Psychology, Keio University, Tokyo, Japan

OPEN ACCESS

Edited by:
Seiki Konishi, Juntendo University, Japan

Reviewed by:
Junichi Chikazoe, National Institute for Physiological Sciences (NIPS), Japan
Hamid Reza Marateb, Universitat Politecnica de Catalunya, Spain

***Correspondence:**
Hiroki Kurashige
h.kura00@gmail.com

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *09 September 2019*
Accepted: *12 December 2019*
Published: *10 January 2020*

Citation:
Kurashige H, Kaneko J, Yamashita Y, Osu R, Otaka Y, Hanakawa T, Honda M and Kawabata H (2020) Revealing Relationships Among Cognitive Functions Using Functional Connectivity and a Large-Scale Meta-Analysis Database. Front. Hum. Neurosci. 13:457. doi: 10.3389/fnhum.2019.00457

To characterize each cognitive function *per se* and to understand the brain as an aggregate of those functions, it is vital to relate dozens of these functions to each other. Knowledge about the relationships among cognitive functions is informative not only for basic neuroscientific research but also for clinical applications and developments of brain-inspired artificial intelligence. In the present study, we propose an exhaustive data mining approach to reveal relationships among cognitive functions based on functional brain mapping and network analysis. We began our analysis with 109 pseudo-activation maps (cognitive function maps; CFM) that were reconstructed from a functional magnetic resonance imaging meta-analysis database, each of which corresponds to one of 109 cognitive functions such as 'emotion,' 'attention,' 'episodic memory,' etc. Based on the resting-state functional connectivity between the CFMs, we mapped the cognitive functions onto a two-dimensional space where the relevant functions were located close to each other, which provided a rough picture of the brain as an aggregate of cognitive functions. Then, we conducted so-called conceptual analysis of cognitive functions using clustering of voxels in each CFM connected to the other 108 CFMs with various strengths. As a result, a CFM for each cognitive function was subdivided into several parts, each of which is strongly associated with some CFMs for a subset of the other cognitive functions, which brought in sub-concepts (i.e., sub-functions) of the cognitive function. Moreover, we conducted network analysis for the network whose nodes were parcels derived from whole-brain parcellation based on the whole-brain voxel-to-CFM resting-state functional connectivities. Since each parcel is characterized by associations with the 109 cognitive functions, network analyses using them are expected to inform about relationships between cognitive and network characteristics. Indeed, we found that informational diversities of interaction between parcels and densities of local connectivity were dependent on the kinds of associated functions. In addition, we identified the homogeneous and inhomogeneous network communities

about the associated functions. Altogether, we suggested the effectiveness of our approach in which we fused the large-scale meta-analysis of functional brain mapping with the methods of network neuroscience to investigate the relationships among cognitive functions.

Keywords: human brain, fMRI, meta-analysis database, functional connectivity, network analysis, data mining, machine learning

INTRODUCTION

Ones of main missions of cognitive neuroscience and psychology is to understand each cognitive function *per se* and to understand the human brain as an aggregate of cognitive functions. To this end, it is vital to relate dozens of cognitive functions, which will provide integrated views for the entire cognition in the human brain and will enable to characterize each cognitive function by relating it to others. Such understanding will provide testable hypotheses for the cognitive neuroscience and psychology communities. In addition, it will give the artificial intelligence community guidelines and ideas to develop novel brain-inspired AI algorithms (Hassabis et al., 2017).

Several efforts in psychology have been conducted to reveal hidden relationships among cognitive functions. Developing atlases and/or ontologies for psychological concepts is one of these endeavors to do so (Price and Friston, 2005; Poldrack et al., 2011; Turner and Laird, 2012; Poldrack and Yarkoni, 2016). Using such ontological data has been shown to be efficient for probing the neural bases of cognitive functions (Varoquaux et al., 2018). Therefore, we consider that building atlases and ontological databases for psychological constructs are promising approaches. However, currently existing atlases and ontological databases are highly conceptual but not sufficiently empirical, which means that most of the relationships are proposed based on the 'common senses' in psychology. It also may lead to missing many meaningful relationships latent in the experimental data which have become big data nowadays. Another effort is to compare cognitive concepts (or psychological constructs) with each other by trying to identify relationships in idiosyncratic features or performances in several cognitive tasks (Beaty et al., 2014; Chuderski and Jastrzêbski, 2018; Eisenberg et al., 2019; Fuhrmann et al., 2019; Rey-Mermet et al., 2019) as well as by investigating overlaps in neural substrates using neuroimaging and neuropsychological methods (Hassabis et al., 2007; Mullally and Maguire, 2014; Woolgar et al., 2018; Brandl et al., 2019; Jonikaitis and Moore, 2019). While these approaches provide insights based on empirical facts, completing such low-profile tasks exhaustively is challenging.

The magnitude of such exhaustive explorations of common or dissociated neural bases among many cognitive functions may dampen the willingness of identification of relationships among them. However, leveraging neuroscientific knowledge is still expected to be effective to our aim because the cognitive functions that overlapping brain regions are responsible for should be interrelated. Additionally, we also consider that the cognitive functions that connected brain regions are responsible for should be interrelated. Therefore, the use of large-scale meta-analysis databases with knowledge about network topology of the brain is essential to find relationships among cognitive functions to characterize each function and the entire cognition in the brain.

BrainMap (Laird et al., 2005, 2011a; Laird, 2009) and Neurosynth (Yarkoni et al., 2011; Poldrack et al., 2012) are databases specialized toward linking cognitive functions to brain regions. The former is a manually constructed database and includes activation coordinates and ontological data (e.g., behavioral domain, task paradigm, and stimulus modality) reported in fMRI studies. The latter is an automated database including activation coordinates and terms extracted from fMRI studies. We can reconstruct pseudo-activation patterns underlying the reports in each study using the stored activation coordinates. Therefore, we are able to relate cognitive functions investigated in the study to the (pseudo-)activation patterns. For instance, the BrainMap's team proposed an approach to provide interpretations of independent components of brain activity based on the cognitive functions (Smith et al., 2009; Laird et al., 2011b; Ray et al., 2013). In another instance, brain parcellation related to cognitive functions was performed by applying decoding based on the cognitive data in BrainMap to parcels identified using connectivity data from the Human Connectome Project database (Fan et al., 2016). In addition, a Bayesian topic model that relates components of cognitive functions to well-localized brain regions was developed (Rubin et al., 2017). This enables decoding of functionality, expressed as rich text information, from any pattern of brain activity. The approaches using BrainMap or Neurosynth are effective for identifying functionalities of sub-divided brain areas, such as the temporoparietal junction (Bzdok et al., 2013), the dorsolateral prefrontal cortex (Cieslik et al., 2013), the insula (Chang et al., 2013), the striatum (Pauli et al., 2016), and the medial frontal cortex (de la Vega et al., 2016). More generally, we can construct pseudo-activation maps corresponding to various cognitive functions (Yarkoni et al., 2011). Hereafter, we term such a pseudo-activation map cognitive function map (CFM).

Here, we explore the relationships among dozens of cognitive functions on the basis of two simple assumptions: (1) cognitive functions that overlapping brain regions are responsible for should be interrelated, and (2) cognitive functions that connected brain regions are responsible for should be also interrelated. To this end, we analyze the CFMs derived from the meta-analysis database with resting-state functional connectivity (RSFC). Therefore, we consider the relationships among cognitive functions from a network neuroscience perspective, which is the subfield of neuroscience to reveal complex but well-organized interdependencies among brain regions using the methods of

network analysis (Sporns and Kötter, 2004; Sporns et al., 2007; Bassett et al., 2008; Hagmann et al., 2008; van den Heuvel et al., 2008; Bullmore and Sporns, 2009, 2012; Power et al., 2011, 2013; van den Heuvel and Sporns, 2011; Zuo et al., 2012; Bertolero et al., 2015; Fornito et al., 2016; Nigam et al., 2016; Bassett and Sporns, 2017; Hwang et al., 2017). We use these methods to reveal relationships among cognitive functions.

In the present study, we analyzed 109 cognitive functions from the viewpoints of connectivity and network analysis using RSFC. First, we provide a comprehensive view of those cognitive functions by constructing relational mapping among them based on the distances quantified as the strengths of RSFCs between the CFMs. This facilitates understanding the brain as a relational network among cognitive functions. Then, we conducted so-called conceptual analysis (in philosophy) of each cognitive function by sub-dividing corresponding CFM on the basis of the connectivity between each voxel within the CFM and the other CFMs, resulting in decomposition of the concept of the function into several sub-concepts. Next, by applying clustering analysis to the whole-brain voxel-to-CFM RSFC, we constructed a whole-brain parcellation where each parcel is labeled with a vector whose components are the degrees of associations to the 109 cognitive functions. Then, by applying matrix factorization to the matrix constructed by concatenating these vectors, we identified six cognitive factors, including 'concept processing,' 'action and expression,' 'vision and attention,' 'executive function,' 'value and judgment,' and 'memory.' Each parcel had degrees of contributions with those factors. Using methods of network analysis to characterize the network consisting of the parcels, we quantified the diversity of the information sources/receivers for the six factors, identified three densely connected subnetworks which are specialized for 'concept processing,' 'action and expression,' and 'vision and attention,' and found (un-)uniformity of factors associated with the parcels within each network community.

The goals of our research are to exhaustively reveal relationships among cognitive functions and relationships between cognitive functions and network characteristics in the brain. Although several previous studies partially suggested such relationships by focusing on some part of the cognitive functions, to the best of our knowledge, there has been no exhaustive effort to those subjects, at least explicitly. Therefore, the main contribution of the present study is, firstly, to provide promising ways to construct comprehensive knowledge of organizations of dozens of cognitive functions as exhaustively as possible. Moreover, we also contribute to providing hopeful ways to reveal relationships between dozens of cognitive functions and network characteristics in the brain. Indeed, we found several new insights into the relationships among cognitive functions and the relationships between cognitive functions and network characteristics. These were achieved by the fusion of large-scale meta-analysis of studies of functional brain mapping and methods in the network analysis.

Taken together, we suggest the effectiveness of our approach in which we fused the large-scale meta-analysis of a task-based fMRI database with network neuroscience approaches to investigate the relationships among cognitive functions to understand each cognitive function *per se* and the human brain as a relational system consisting of cognitive functions.

MATERIALS AND METHODS

Subjects

Fifty-two subjects (21 women) without a history of neurological or psychiatric diseases participated in this study. The mean ages of the male and female subjects were 21.5 and 22.3 years (standard deviation, 1.27 and 6.94 years), respectively. All subjects were right-handed. They had a normal or corrected-to-normal vision. We did not use any power analysis to determine the sample size but decided the size by reference to previous resting-state fMRI studies (e.g., Fox et al., 2005; Honey et al., 2009; Smith et al., 2009). To recruit participants, we mainly used announcements through Web sites (including SNS) and snowball sampling.

The study was performed in accordance with the recommendations of the institutional ethics committee of the National Center of Neurology and Psychiatry (NCNP), with written informed consent from all subjects, in accordance with the Declaration of Helsinki. The institutional ethics committee of the NCNP approved the study protocol (Approval No. A2014-072).

MRI Acquisition

We used a 3T MRI scanner (Trio, Siemens Medical Solutions, Erlangen, Germany) with an 8-channel head coil for all measurements. Structural images were acquired using a T1-weighted 3D magnetization-prepared-rapid-gradient-echo sequence. The parameters used were: flip angle = 8°, voxel size = 1 mm isotropic, TR = 2000 ms, TI = 990 ms, TE = 4.38 ms, and number of voxels = 208 × 256 × 208. Functional images were acquired with a T2*-weighted echo-planar imaging sequence. The parameters used were: flip angle = 90°, voxel size = 3 mm (isotropic, with no slice gap), TR = 3000 ms, TE = 30 ms, and number of voxels = 64 × 64 × 44. The slices were acquired in interleaved order.

Resting-State fMRI

We acquired 148 volumes of images. As TR was 3 s, the total acquisition time was approximately 7.4 min. During imaging, a fixation point centered on a gray background was presented. We instructed the subjects to look at the fixation point and to think of nothing in particular.

Preprocessing of MRI Data

We performed the preprocessing mainly using FSL (FMRIB Software Library Version 5.0.6[1]) (Jenkinson et al., 2012). All steps were executed by running commands in FSL from custom-made shell scripts.

First, we applied slice-time correction to functional images using the slicetimer command. Next, we conducted head motion correction using the mcflirt command (Jenkinson et al., 2002) with the '-stages 4 -sinc_final -meanvol -mats -plots' option.

[1]http://www.fmrib.ox.ac.uk/fsl/

Afterward, we applied the bet command (Smith, 2002) to structural images to extract the brain regions from whole images. For these structural images, the flirt command (Jenkinson and Smith, 2001; Jenkinson et al., 2002; Greve and Fischl, 2009) was executed with the MNI152_T1_2mm_brain template as a reference. In this step, we used six degrees of freedom, resulting in rigid-body transformation. Therefore, we executed this step only for alignment and changing the resolution to 2 mm. Then, we applied the bet command to the mean functional image and obtained registration parameters of the image to the 2-mm-resolution structural image using the flirt command. Using these parameters, we registered all functional images to the 2 mm-resolution structural image, resulting in 2-mm-resolution functional images. Next, we obtained non-linear transformation parameters by applying the fnirt command to the 2-mm-resolution structural image, with the MNI152_T1_2mm template as a reference. Then, we transformed the 2-mm-resolution functional images using the applywarp command with the non-linear transformation parameters. This yielded 2-mm-resolution functional images that were standardized into the Montreal Neurological Institute (MNI) 152 space. Additionally, we masked these functional images with the regions of the MNI152 standard brain and smoothed them with a 5-mm full-width at half-maximum. These functional images were used in the following analyses.

Additionally, the structural image was standardized into the 1-mm-resolution MNI 152 space followed by the recon-all process in Freesurfer (version 5.3.0[2]). This yielded cortical and subcortical atlases (Fischl et al., 2002; Desikan et al., 2006) standardized into the 1-mm-resolution MNI 152 space.

In the analyses for the resting-state fMRI shown in the following subsections, we excluded subjects whose translational head motions were 1 mm or more or whose rotational head motions were 1° or more, since head motion severely affects the inference of RSFC (Power et al., 2012; van Dijk et al., 2012). Our criterion is more stringent compared with the conventional criteria from previous studies (Guo et al., 2012; Jackson et al., 2016; Liu et al., 2016; Zhu et al., 2017). According to the criterion, we excluded twenty-five subjects We did not adopt any other criterion for excluding data.

Whole-Brain Anatomical Atlas

To construct a whole-brain anatomical atlas, we used the output files of the recon-all process in Freesurfer. As described above, the input file for the process was an individual structural image standardized into the MNI152 space. Therefore, the output file provided the whole-brain atlas for each subject standardized into the MNI152 space. In this atlas, each voxel is labeled with an intensity to specify the anatomical area according to the Freesurfer convention.

We decomposed the whole-brain atlas for each subject to the anatomical regions. For each region, we aggregated the atlases for all subjects into one average atlas by the following method. First, for each voxel, we counted the number of subjects whose individual atlases for the region included the voxel and assigned it to the voxel. Then, we binarized the resulting image with a threshold of the number of subjects for the inclusion of voxels into the aggregated atlas, which made the number of voxels in the image closest to the mean of the number of voxels composing the region across the subjects. This provided an average atlas across the subjects for the anatomical regions. Finally, we merged these average atlases into one whole-brain anatomical atlas on the MNI152 standardized brain. In this whole-brain atlas, each voxel is labeled with the intensity indicating the corresponding anatomical region in a manner following the Freesurfer convention.

Construction of Pseudo-Activation Maps

We constructed a pseudo-activation map for each cognitive function. To this end, we followed the method based on χ^2 statistics described previously (Yarkoni et al., 2011). We will give an in-depth explanation of the procedure in the remainder of this section. In the procedure, we used version 0.4 of Neurosynth data downloaded from the Neurosynth page on GitHub[3].

First, for the articles registered in Neurosynth data, we obtained titles, keywords, and abstracts by accessing PubMed[4] using the Entrez Programing Utilities (E-utilities) API[5] executed from the Biopython module (Cock et al., 2009) in Python. Then, we counted the appearances of cognitive concepts in the title, keywords, and abstract for each article. As for the cognitive terms considered in this study, we prepared 702 concepts. Of these, 692 were extracted from the list named 'concepts' in Cognitive Atlas (Poldrack et al., 2011). The extraction date was 8/18/2014. We added ten cognitive terms. The cognitive terms that we considered are listed in **Supplementary Table S1**.

We considered a cognitive term to be present in an article if the term appeared one or more times per 100 words in the text merged from the title, keywords, and abstract of the article. We included only the cognitive terms that appeared in ten or more articles in the following analyses. Additionally, we discarded the terms that are used as general words in neuroscience literature, such as 'focus' and 'strength.' Thus, we selected the 121 cognitive terms shown in **Supplementary Table S2** as the targets to be considered in this study.

Then, we reconstructed the binary activation map on the 2-mm-resolution MNI 152 brain for each article registered in Neurosynth data by the following steps. First, we transformed the coordinates reported in the Talairach brain into the MNI brain using icbm2tal transform (Lancaster et al., 2007). Then, we assigned the number '1' to the voxels located within 10 mm of the registered coordinates and the number '0' to the other voxels. Based on these binary activation maps, we calculated the χ^2 statistics for each cognitive term, in which we compared the appearance of the term and activation of the voxel. Additionally, we calculated the ϕ coefficients corresponding to the χ^2 statistics. Thus, we obtained χ^2 and ϕ maps for each cognitive term. For convenience, in the following statistical test, these maps were masked by voxels that were

[2]https://surfer.nmr.mgh.harvard.edu/

[3]https://github.com/neurosynth

[4]https://www.ncbi.nlm.nih.gov/pubmed/

[5]https://www.ncbi.nlm.nih.gov/books/NBK25501/

activated in 3% or more articles, which reduced the number of voxels to be tested.

Based on the χ^2 test using the χ^2 map, we constructed a pseudo-activation map for each cognitive function in the following manner. We executed multiple-test corrections using the Benjamini–Hochberg procedure for controlling the false discovery rate (Benjamini and Hochberg, 1995) with $q^* = 0.05$. This yielded a significant mask for each cognitive function. Additionally, we constructed a mask for each cognitive function where the positive values of ϕ coefficients indicated a positive correlation. Applying these masks to χ^2 or ϕ maps, we obtained pseudo-activation maps where the pattern of significant positive activation induced with the cognitive function is expressed. We call these pseudo-activation maps CFMs. In the present study, 109 CFMs had significant voxels. Therefore, we focused on these 109 cognitive functions (**Supplementary Table S3**).

Two-Dimensional Embedding of Cognitive Concepts Based on the CFM-to-CFM RSFC Matrix

We constructed a two-dimensional relational map among the 109 cognitive functions based on the time-series data of blood-oxygen-level-dependent (BOLD) signals for each CFM. First, we extracted the time-series data of BOLD signals of resting-state fMRI for each voxel in the whole-brain mask. To reduce artifacts due to motion and signal drift, six head motion parameters and six differential values of head motion plus the linear trend and constant component were regressed out. Then, a 0.009–0.08 Hz band-pass filter was applied to remove the putative non-neuronal signals according to previous reports (Biswal et al., 1995; Cordes et al., 2001; Lu et al., 2007; Zuo et al., 2010). We used the 5th-order Butterworth digital filter. This filter was applied in forward and backward. We confirmed that further increase of the order led to little change the resulting waveform. In addition, the average signals of the gray matter region, white matter region, and ventricles were regressed out. Those data were transformed to Z-scores by each voxel to erase the intensity bias between the voxels. For all voxels for all subjects, the maximum and minimums Z-scores were 5.32 and −4.89, respectively. By applying the Kolmogorov–Smirnov test to each voxel of each subject, we found that 0.38% voxels were judged as non-normal distributions. Such a preprocessing flow was used also in the further analyses described below.

Then, for each subject, we obtained the mean signal for each CFM by averaging the signals across voxels in the CFM. We calculated the correlation matrices between signals of CFMs and averaged them across the subjects, resulting in the CFM-to-CFM RSFC matrix. By shifting and scaling the RSFC values, we obtained the CFM-to-CFM similarity matrix in which the minimum and maximum values were 0 and 1, respectively. Applying spectral clustering (see the next paragraph) to the similarity matrix, we identified clusters of cognitive functions. In this step, we determined the number of clusters as the value corresponding to the maximum of silhouette coefficients (Rousseeuw, 1987) up to 12 (**Supplementary Figure S1**).

The reason why we used the spectral clustering to identify the clusters of cognitive functions is that our problem in this analysis was based on the similarity matrix (not on the feature vectors). For convenience to the readers, we give a brief introduction to spectral clustering (von Luxburg, 2007). The procedure of the spectral clustering consists of two steps. The first step is to embed data into a representational space. In this space, coordinates (or representations) of the data are determined to minimize a loss defined with the similarity matrix and the coordinates. This minimization problem is reduced to the eigenvalue problem. Except for parameters used in the numerical calculus to solve the eigenvalue problem, the parameter that we need to set is the dimension of the representational space that is equal to the number of eigenvectors that we consider. Throughout the present study, we set this value to the same as the number of clusters. The second step is to cluster the data based on the coordinates in the representational space. In this step, we need to determine the way to cluster. Here, we used k-means clustering.

Finally, we applied multidimensional scaling (Borg and Groenen, 1997) to the CFM-to-CFM RSFC matrix using the scikit-learn module in Python and obtained the relational map that involves two-dimensional embedding of the 109 cognitive functions, in which the well-correlated pairs of cognitive functions were located as closely as possible. To check a distortion caused by the embedding, we calculated the stress that is defined as the difference between given dissimilarities and distances in the embedding space and is the value to be minimized in the multidimensional scaling.

Subparcellation of CFMs

For each cognitive function and subject, the resting-state fMRI BOLD signals of the voxels in the corresponding CFM were extracted and preprocessed in the same manner as described in the previous sections. Now we focus on a CFM that we term target CFM. We calculated the correlation values between the processed signals of all voxels in the target CFM and the mean signals obtained from the other CFMs by averaging signals across the voxels belonging to the CFMs. These correlation values were averaged across the subjects. Thus, we obtained the target CFM voxel-to-CFM RSFC matrix. For instance, if we express the number of voxels in the 'emotion' CFM as N (emotion), the RSFC matrix has a dimension of N(emotion) × 108, since we considered 109 cognitive functions.

Then, we executed principal component analysis for dimensionality reduction. We determined the number of principal components required to explain 95% of the total variance. Finally, we applied k-means clustering to the resulting data, in which we set the number of clusters as five.

Whole-Brain Parcellation Based on Voxel-to-CFM Functional Connectivity

We conducted whole-brain parcellation based on RSFC. First, we extracted the time series data of BOLD signals of resting-state fMRI for each voxel in the whole-brain mask and calculated the mean signal for each CFM. These procedures were the same as those described above.

From these processed data, we obtained a matrix of voxel-to-CFM RSFC by calculating correlation coefficients between the processed BOLD signals of the 160,296 voxels in the whole-brain mask and the processed BOLD signals of the 109 CFMs for each subject. Then, we transformed them into Fisher's Z values and averaged them across the subjects. We applied inverse Fisher's Z-transform to this and obtained the mean voxel-to-CFM RSFC matrix in which each row was a 109-dimension feature vector for each voxel. To reduce the number of dimensions, we performed principal component analysis by solving the eigenvalue problem for the covariant matrix for the voxels. We determined the number of principal components required to explain 95% of the total variance. Based on this dimension-reduced matrix, we constructed a similarity matrix between voxels, where the similarity was defined as the exponential of the correlation coefficient between two voxels. To consider the spatial constraint, we set the similarity between the voxels that were not neighbored to 0, according to a previous study (Craddock et al., 2012), resulting in a sparse similarity matrix.

To obtain whole-brain parcellation, we applied multiclass spectral clustering (Yu and Shi, 2003) to this similarity matrix using the scikit-learn module in Python with 'discretize' (to use the optimal discretization approach searching a discrete partition closest to the continuous representations to identify data clusters in the representational space identified with spectral embedding) and 'amg' options. (The reason why we adopted the spectral clustering in this analysis was to use the spatial constraint mentioned above.) Since this algorithm requires the similarity matrix to be connected, we randomly chose 500,000 pairs of voxels and assigned them a weak positive value (0.0001). We set the number of clusters to 200 that was determined by reference to several existing atlases (Destrieux et al., 2010; Power et al., 2011; Shen et al., 2013; Baldassano et al., 2015; Fan et al., 2016). This resulted in whole-brain parcellation with 199 parcels. One cluster was discarded because it was empty (no voxel). We assigned each parcel a label vector that was the mean voxel-to-CFM RSFC obtained by averaging voxel-to-CFM RSFCs across the voxels belonging to the parcel, which represents the relatedness between the parcel and the 109 cognitive functions (parcel-to-CFM RSFC matrix).

Dimensionality Reduction Using the Non-negative Matrix Factorization

We applied the non-negative matrix factorization (NMF) (Lee and Seung, 1999, 2001) to the parcel-to-CFM matrix to reduce the dimensionality, which was executed using the NIMFA module (Žitnik and Zupan, 2012) in Python. Before this process, we set the negative values in the matrix to 0.

The NMF is a method to decompose a non-negative data matrix ($\mathbf{X}$) into a non-negative coefficient matrix ($\mathbf{Y}$) and a non-negative basis matrix ($\mathbf{Z}$). The objective of the NMF is to approximate $\mathbf{X}$ by $\mathbf{YZ}$. Thus, we used the Frobenius norm $||\mathbf{X}-\mathbf{YZ}||_F$ as the cost function and minimized it subject to the $\mathbf{Y} \geq 0$ and $\mathbf{Z} \geq 0$. Our purpose in the dimensionality reduction was to identify well-interpretable low dimensional representations for the parcels. In the preprocessing procedure, we regressed the mean time-course of the gray matter signals out from the BOLD data. Although this is efficient to remove artifacts resulting from biological and equipment factors (Satterthwaite et al., 2013; Power et al., 2014; Li et al., 2019), it is suggested that this procedure tends to cause artifactual negative correlation (Murphy et al., 2009; Weissenbacher et al., 2009). Therefore, to lead better interpretation for the parcels, focusing only on the positive RSFCs is appropriate. Therefore, we chose the NMF as the way for dimensionality reduction.

The number of factors is a key parameter to be predefined in the NMF. A previous study suggests that the inflection point in the decrementing line of residual sum of squares (RSSs) with an increment of the values of the numbers of factors yields the adequate number (Hutchins et al., 2008). We can detect the inflection point as the crossing point between the curved lines fitted to RSSs before and after the point. Therefore, we first calculated the differentials of the RSSs and fitted them to straight lines. We repeated the linear regression and obtained the sums of the squared errors of before-point and after-point lines while changing the point. We determined the inflection point as the point realizing the minimum value of the summed squared error (**Supplementary Figure S2**). Using the value corresponding to the point as the number of factors, we conducted the NMF with singular value decomposition (SVD)-based initialization (Boutsidis and Gallopoulos, 2008).

Since the output vectors constituting the bases were not normalized, we scaled them to generate unit vectors and applied the inverse operation of the scaling to the coefficient matrix to keep the product invariant.

Heat-Diffusion Analysis of Information Sources/Receivers

We extracted the time series data of BOLD signals of resting-state fMRI for each parcel in the whole-brain parcellation obtained above. These data were preprocessed in the same manner described in the previous sections. We calculated correlation coefficients between the processed BOLD signals of the parcels and obtained a parcel-to-parcel RSFC matrix averaged across subjects. We set the negative values and diagonal components in the matrix to 0 and treated it as an adjacency matrix $\mathbf{A}$. In addition, we defined the degree matrix $\mathbf{D}$ in which the diagonal components were $D_{ii} = \sum_j A_{ij}$ and the other components were 0. From these matrices, we defined graph Laplacian matrix $\mathbf{L} = \mathbf{D}-\mathbf{A}$ (Chung, 1997), which is the homolog of the negative Laplacian $-\nabla^2$.

For each NMF factor, we regarded the values of the NMF coefficients as the intensities of the heat sources distributed over the parcels. Based on the heat source distribution, we calculated the steady temperature distribution on the graph whose links were defined by the adjacency matrix $\mathbf{A}$ between parcels as graph nodes in the following manner, according to a procedure developed in network theory (Newman, 2010). In the usual partial differential equations, the temperature diffusion ψ with heat sources f is governed by the equation $\partial\psi/\partial t = D\nabla^2\psi - \beta\psi + \mathbf{f}$, where D is a diffusion coefficient and β is a decay constant. As an analog of this equation for the graph, we obtained

the equation $\partial\boldsymbol{\psi}/\partial t = -D\mathbf{L}\boldsymbol{\psi} - \beta\boldsymbol{\psi} + \mathbf{f}$, where $\mathbf{f}$ is the vector of heat source distribution defined as the NMF coefficient. As we consider the steady state $\partial\boldsymbol{\psi}/\partial t = \mathbf{0}$. Therefore, the steady temperature distribution is $\boldsymbol{\psi} = (D\mathbf{L} + \beta\mathbf{E})^{-1}\mathbf{f}$, where $\mathbf{E}$ is an identity matrix. We set $D = 1$ and $\beta = 1$ in the main analysis.

The temperature distribution was calculated for each NMF factor. The temperature of the parcels for each NMF factor represents a degree of information conveyed from the factor. Thus, each parcel has a vector of temperatures of the NMF factors. From the vector, we calculated the Gini coefficients that represent disparity of conveyed information among the NMF factors. If a parcel receives information from only one factor, the value of the Gini coefficient becomes 1. Conversely, if a parcel receives information from all factors uniformly, the value becomes 0.

We defined the Gini coefficient for each cognitive function as the mean of 10 Gini coefficients of parcels whose parcel-to-CFM RSFCs for the function were the top ten values. In other words, we averaged the ten Gini coefficients of parcels that were the most related to the cognitive function and considered the resulting mean as the Gini coefficient for the function.

In an additional analysis, we investigated the effects of the parameter values. Since the result is dependent only on the ratio of D and β, only D was varied and β was fixed ($\beta = 1$). Here, we compared the Gini coefficients between NMF factors. For each factor, we calculated an inner product between the vector of the Gini coefficients for cognitive functions defined above and the vector of the corresponding NMF basis that was normalized to make the summation one. We call this inner product weighted sum of the Gini coefficients. Intuitively, the weighted sum of the Gini coefficients expresses the mean of the Gini coefficients for the cognitive functions assigned to the factor.

Local Density Identification in the Parcel-to-Parcel Network Using Clique Percolation

In this analysis, we first created the parcel-to-parcel network by defining the connectivity among parcels by thresholding the adjacency matrix **A** with 0.3. Then, we applied the clique percolation method (Palla et al., 2005) to this network to investigate the local densities of connectivity in this network using the networkX Python module. In graph theory, K-clique implies the fully connected subgraph consisting of K nodes. In the clique percolation method, first, K-cliques are identified. Then, pairs of K-cliques are connected to form a cluster if they share a (K-1)-clique. Furthermore, if the cluster shares a (K-1)-clique with another K-clique, it is assimilated to the cluster. This process is iteratively executed. When we set K to a large value, the resulting cluster becomes a densely connected subgraph.

Community Analysis on the Parcel-to-Parcel RSFC Matrix

By shifting and scaling the values in the parcel-to-parcel RSFC matrix, we first obtained the parcel-to-parcel similarity matrix in which the minimum and maximum values were 0 and 1, respectively. To identify the community structure in the parcels based on the similarity matrix, we applied spectral clustering to the parcel-to-parcel similarity matrix using the scikit-learn module in Python. As is the case with the clustering of cognitive functions, we used the k-means method to identify data clusters in the representational space identified with spectral embedding. The number of communities was set at the value maximizing the silhouette coefficients (Rousseeuw, 1987) up to 20 (**Supplementary Figure S3**). The other parameters were set to the default values.

Reliability Check of RSFC Matrices

Since the analyses described in the above subsections were basically based on the RSFC matrices defined as the correlation matrices, checking the reliabilities of the estimations is worthful to evaluate the stabilities of the results. Especially, we should be careful about the possible instabilities that might be caused by the smaller data size compared to the data stored in the recently developing large-scale databases such as the Human Connectome Project database (Smith et al., 2013; Van Essen et al., 2013). To this end, we calculated the standard errors of means (SEMs) of the RSFCs. Accordingly, we observed the small levels of the values ($\sim$0.035) compared to the absolute RSFC values (**Supplementary Figure S4**), which means that the effects of the instabilities caused by the small data size were substantially limited.

We also conducted a correlation analysis between the RSFCs estimated from the present data and the Human Connectome Project data (**Supplementary Figure S5**). We used only 706 of about 2000 data in S500 dataset because of resource limitation. The preprocessing pipeline was the same as the one explained above. The correlation coefficients are acceptable (0.94 for CFM-to-CFM RSFCs, 0.84 for voxel-to-CFM RSFCs, and 0.58 for parcel-to-parcel RSFCs). Again, those results suggest that the small size of the present data affected the results limitedly.

RESULTS

Relational Mapping for Cognitive Functions

In the present study, we aimed to elucidate the relationships among the cognitive functions in the human brain. To obtain a comprehensive overview of the human cognition, a visualization of the whole picture representing the relationships among cognitive functions is required. To this end, we began our analysis with the 109 CFMs which were reconstructed as pseudo-activation maps corresponding to 109 cognitive functions (**Figure 1A**). By applying multidimensional scaling to the CFM-to-CFM RSFC matrix (**Figure 1B** and **Supplementary Data S1**), we provided a relational mapping that involved two-dimensional embedding of the cognitive functions, in which the closely related cognitive functions were located close to each other (**Figure 1C**). In addition, we identified six clusters of cognitive functions (cognitive clusters) using the spectral clustering method with the silhouette coefficients (**Figure 1C** and **Table 1**). Roughly, the red-purple cluster included 'self and others'-related functions, the blue cluster included 'executive function'-related functions, the orange cluster included 'language'-related functions, the

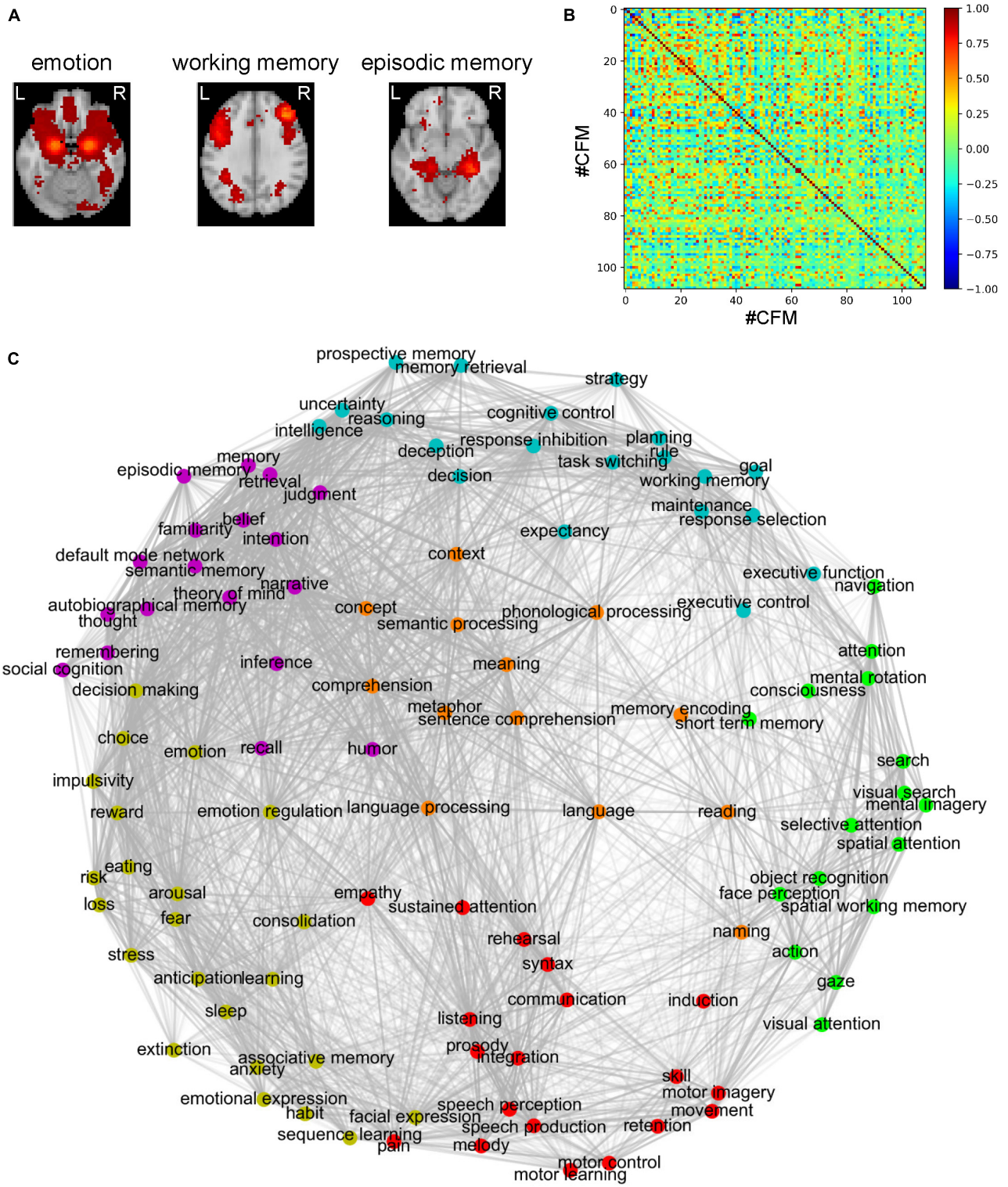


FIGURE 1 | Relational mapping and clustering of cognitive functions based on the cognitive function map (CFM)-to-CFM resting-state functional connectivities (RSFCs). **(A)** Examples of the reconstructed CFMs. **(B)** The CFM-to-CFM RSFC matrix. **(C)** Relational mapping of cognitive functions. The locations of cognitive functions in the two-dimensional plane were determined by applying multidimensional scaling to the CFM-to-CFM RSFC matrix. Furthermore, spectral clustering was applied and resulted in six clusters of cognitive functions, in which, roughly, the red-purple cluster included 'self and others'-related functions, the blue cluster included 'executive function'-related functions, the orange cluster included 'language'-related functions, the yellowish-green cluster included 'value and judgment'-related functions, the red cluster included 'action and expression'-related functions, and the green cluster included 'vision and attention'-related functions.

TABLE 1 | Clustering of the cognitive functions based on the CFM-to-CFM RSFC matrix.

Action and expression	Vision and attention	Value and judgment	Self and others	Executive function	Language
Movement	Attention	Emotion	Memory	Working memory	Language
Pain	Action	Reward	Retrieval	Decision	Reading
Integration	Gaze	Learning	Judgment	Cognitive control	Context
Skill	Spatial attention	Risk	Intention	Response inhibition	Meaning
Empathy	Selective attention	Fear	Recall	Goal	Comprehension
Listening	Search	Anxiety	Episodic memory	Rule	Concept
Motor imagery	Navigation	Decision-making	Default mode network	Reasoning	Naming
Prosody	Short-term memory	Stress	Familiarity	Maintenance	Semantic processing
Speech perception	Mental rotation	Loss	Social cognition	Planning	Metaphor
Communication	Consciousness	Choice	Inference	Executive function	Memory encoding
Sustained attention	Spatial working memory	Anticipation	Belief	Uncertainty	Language processing
Motor control	Visual search	Sleep	Thought	Deception	Phonological processing
Retention	Visual attention	Facial expression	Theory of mind	Task switching	Sentence comprehension
Rehearsal	Mental imagery	Arousal	Semantic memory	Strategy	
Syntax	Face perception	Emotion regulation	Narrative	Response selection	
Induction	Object recognition	Extinction	Autobiographical memory	Executive control	
Speech production		Impulsivity	Humor	Intelligence	
Motor learning		Habit	Remembering	Memory retrieval	
Melody		Eating		Expectancy	
		Consolidation		Prospective memory	
		Sequence learning			
		Associative memory			
		Emotional expression			

CFM, cognitive function maps; RSFC, resting-state functional connectivity.

yellowish-green cluster included 'value and judgment'-related functions, the red cluster included 'action and expression'-related functions, and the green cluster included 'vision and attention'-related functions.

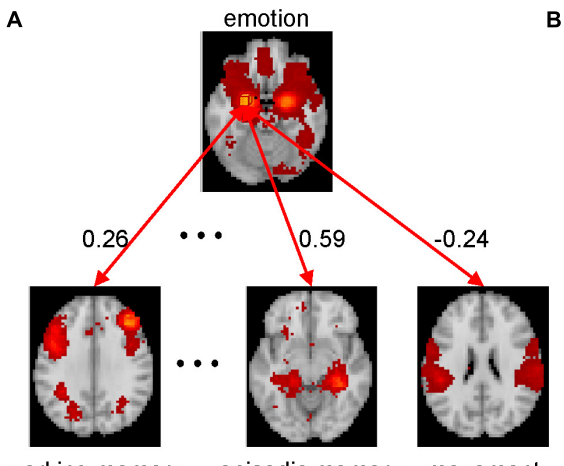
To check a distortion caused by the embedding, for each embedding dimension up to ten, we calculated the stress that is an index quantifying the deviation of distances in the embedding space from the distances defined based on the similarity matrix. The decline of stress is shown as the scree plot in **Supplementary Figure S6**. According to the scree criterion, an optimal dimension seems to be four. Although the two-dimensional mapping has good readability, this means that it was somewhat distorted and could not exactly express the strengths of the RSFCs between the CFMs. Therefore, we also provide figures that are similar to **Figure 1C** but show the positive and negative strengths of RSFCs using red and blue colors, respectively (**Supplementary Figure S7**).

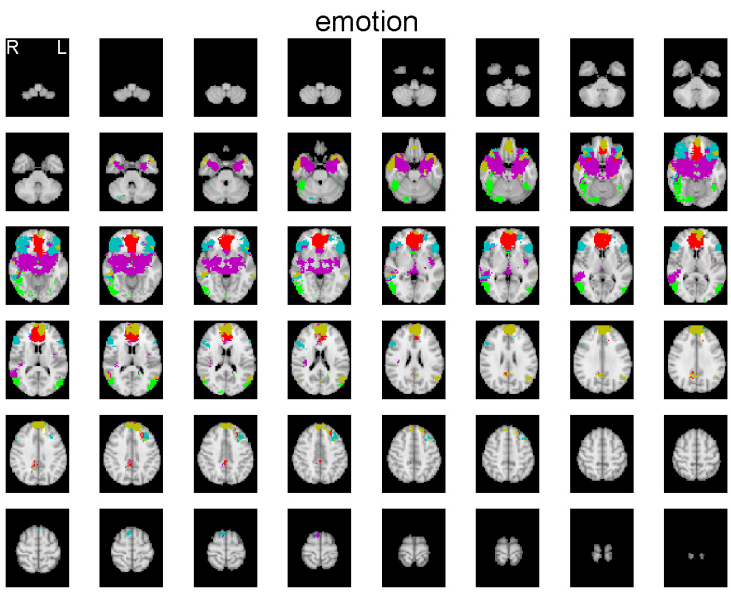
RSFC-Based Conceptual Analysis of Cognitive Functions

One of the bottlenecks preventing us from understanding information processing during cognitive functions is that we do not have sufficient in-depth knowledge of the concepts of these cognitive functions. Therefore, we require so-called conceptual analysis of the cognitive functions (based not on philosophical deliberation but on neuroscientific evidence) to elucidate their deeper meanings. Here, we propose a method of conceptual analysis based on the voxel-to-CFM RSFCs (**Figure 2**). In this method, first, we selected a cognitive function (e.g., 'emotion') and the corresponding CFM. For all voxels within the selected CFM and the 108 remaining CFMs, a voxel-to-CFM RSFC matrix was constructed (**Figure 2A**). Then, we applied k-means clustering to the matrix and subdivided the CFM for the cognitive function into five clusters (**Figure 2B**). Finally, each cluster was related to the 108 remaining cognitive functions based on cluster-to-CFM RSFCs.

As examples, the results for 'emotion,' 'prospective memory,' and 'thought' are shown in **Figure 2C**. The subdivision of the 'emotion'-corresponding CFM suggests that 'emotion' is constructed of the subfunctions related to decision-making (cluster 1), vision (cluster 2), self and others' minds (cluster 3), fear (cluster 4), and comprehension of abstract meanings (cluster 5). The subdivision of the 'prospective memory'-corresponding CFM suggests that 'prospective memory' is constructed of the subfunctions related to memory (cluster 1), intelligent decision (cluster 2), motion (cluster 3), emotional decision (cluster 4), and executive function (cluster 5). The subdivision of the 'thought'-corresponding CFM suggests that 'thought' is constructed of the subfunctions related to self and others' minds (cluster 1), imaginary navigation (cluster 2), logical intention and intelligence (cluster 3), emotional decision (cluster 4), and memory (cluster 5).

The results for all cognitive functions are provided in **Supplementary Data S2**. In this analysis, we set the numbers of clusters identical (i.e., five) across all CFMs by considering interpretability. On the other hand, showing results from the clustering in which the numbers of clusters were determined





C

emotion

cluster 1	cluster 2	cluster 3	cluster 4	cluster 5
decision making 0.464	gaze 0.337	theory of mind 0.379	fear 0.221	context 0.304
choice 0.457	face perception 0.333	narrative 0.364	anxiety 0.21	metaphor 0.264
thought 0.402	object recognition 0.322	belief 0.363	facial expression 0.198	decision 0.241
reward 0.389	spatial attention 0.322	inference 0.36	emotional expression 0.187	meaning 0.235
arousal 0.377	visual search 0.255	judgment 0.353	arousal 0.176	emotion regulation 0.231
default mode network 0.37	naming 0.253	familiarity 0.319	anticipation 0.149	sentence comprehension 0.23
impulsivity 0.333	mental imagery 0.211	autobiographical memory 0.318	habit 0.14	language 0.228
stress 0.313	search 0.191	default mode network 0.313	pain 0.133	phonological processing 0.205
fear 0.279	selective attention 0.183	intention 0.311	reward 0.132	cognitive control 0.204
familiarity 0.266	reading 0.18	concept 0.307	emotion regulation 0.132	deception 0.198

prospective memory

cluster 1	cluster 2	cluster 3	cluster 4	cluster 5
default mode network 0.457	intelligence 0.342	action 0.236	decision making 0.381	expectancy 0.444
memory retrieval 0.443	reasoning 0.284	movement 0.216	default mode network 0.377	cognitive control 0.406
retrieval 0.431	uncertainty 0.282	spatial working memory 0.215	choice 0.372	rule 0.392
reasoning 0.416	decision making 0.282	motor imagery 0.213	thought 0.334	response inhibition 0.368
episodic memory 0.37	cognitive control 0.27	skill 0.167	emotion 0.324	intelligence 0.339
autobiographical memory 0.365	choice 0.265	mental rotation 0.163	impulsivity 0.32	working memory 0.33
belief 0.329	expectancy 0.241	maintenance 0.154	intelligence 0.305	response selection 0.297
memory 0.327	memory retrieval 0.232	goal 0.151	reward 0.301	planning 0.291
thought 0.306	default mode network 0.221	planning 0.138	arousal 0.283	deception 0.26
familiarity 0.301	retrieval 0.218	attention 0.122	belief 0.282	decision 0.255

thought

cluster 1	cluster 2	cluster 3	cluster 4	cluster 5
autobiographical memory 0.464	navigation 0.205	intention 0.258	choice 0.599	episodic memory 0.487
default mode network 0.449	mental imagery 0.181	uncertainty 0.256	decision making 0.596	remembering 0.448
familiarity 0.432	visual attention 0.092	intelligence 0.244	default mode network 0.476	retrieval 0.378
theory of mind 0.391	visual search 0.084	judgment 0.235	reward 0.475	navigation 0.37
belief 0.359	spatial attention 0.082	reasoning 0.223	arousal 0.47	memory 0.355
emotion 0.358	remembering 0.08	context 0.217	emotion 0.454	autobiographical memory 0.338
narrative 0.346	episodic memory 0.078	deception 0.215	stress 0.417	default mode network 0.332
judgment 0.345	search 0.071	narrative 0.213	impulsivity 0.403	familiarity 0.328
remembering 0.336	strategy 0.06	decision 0.205	fear 0.362	semantic memory 0.319
memory 0.334	motor control 0.059	belief 0.198	familiarity 0.34	memory retrieval 0.218

FIGURE 2 | Conceptual analysis of cognitive functions based on subdivisions of the cognitive function maps (CFMs). **(A)** Schematic illustration of voxel-to-CFM resting-state functional connectivities (RSFCs) between voxels in a CFM of cognitive function under consideration (e.g., 'emotion') and the other CFMs. **(B)** Example of subdivision of the CFM of 'emotion' obtained by applying k-means clustering to the CFM. In this analysis, the number of clusters was fixed to five. Each color corresponds to each cluster resulting from the subdivision. **(C)** Examples of subdivision-based conceptual analyses for 'emotion' (upper), 'prospective memory' (middle), and 'thought' (lower). The cluster-to-CFM RSFCs are shown with the names of corresponding cognitive functions. For each cluster, the cluster-to-CFM RSFCs are defined as the mean values of the voxel-to-CFM RSFCs for the voxels belonging to the cluster.

based on the silhouette coefficients are beneficial. Therefore, we provide these results in which the numbers of clusters were determined based on the silhouette coefficients (up to twelve clusters) in **Supplementary Data S3**.

The nifti-formatted images of the subdivided CFMs will be downloadable from the authors' web page.

Cognitive Function-Based Whole-Brain Parcellation

Network analyses using brain parcels that are associated with cognitive functions as network nodes are promising to offer insights into the characteristics of each function *per se* and the relationships among those functions. To construct such parcels, we administered a novel whole-brain parcellation method in which voxels were assembled to one of 199 clusters (or parcels) by applying spectral clustering to the voxel-to-CFM RSFC matrix (**Figures 3A–C**). Each resulting parcel was characterized by its relatedness with the 109 cognitive functions (i.e., parcel-to-CFM RSFCs), defined as mean voxel-to-CFM RSFCs over the voxels belonging to the parcel (**Figure 4** and **Supplementary Table S4**). Their links to the anatomical brain regions are provided in **Supplementary Table S5**.

We also show the correspondence between the present parcellation and the Glasser's atlas (Glasser et al., 2016) in **Supplementary Table S6**. We found that the voxels belonging to one parcel in the present parcellation are assigned to several parcels in the Glasser's atlas. This is natural since the number of parcels in the Glasser's atlas is larger than ours. We show the ratios of voxels assigned to the most overlapping region, the second most overlapping region, the third most overlapping region,. . . in **Supplementary Figure S8**. Thirty six percent of the voxels are included in the most overlapping regions in the Glasser's atlas. Up to the fourth most overlapping regions, 84% of the voxels are included in them.

The nifti-formatted CFM and parcellation images will be downloadable from the authors' web page.

Cognitive Factor Identification Based on Dimensionality Reduction Using Non-negative Matrix Factorization

The 109 cognitive functions were not independent of each other. Some functions were highly interrelated, and therefore, had common latent cognitive factors. We believe that all cognitive functions can be characterized by combinations of a few latent cognitive factors. When a group of cognitive functions is commonly dependent on such factors, the parcel-to-CFM RSFCs of members of the group should be similar. Thus, to identify the latent cognitive factors, we applied non-negative matrix factorization (NMF) to the parcel-to-CFM RSFC matrix (**Figure 5A**).

The number of NMF factors was determined to be six according to the evaluation of the residual sum of squares. The top ten components for each basis vector (row vector in the identified NMF basis matrix) with the corresponding cognitive functions are provided in **Table 2**. All components in the bases are shown in **Supplementary Table S7**. The NMF coefficient matrix is shown in **Supplementary Table S8**. We found that these cognitive factors roughly corresponded to 'concept processing' (factor 1), 'action and expression' (factor 2), 'vision and attention' (factor 3), 'executive function' (factor 4), 'value and judgment' (factor 5), and 'memory' (factor 6).

For each factor, the heat map of the NMF coefficients for the corresponding parcels are shown in **Figure 5B**, in which we observe factor-specific spreading patterns. The factor 1-related parcels are located on the left inferior parietal cortex, left superior and middle temporal cortex, left inferior frontal gyrus, and the left superior frontal cortex. The factor 2-related parcels are located on the bilateral sensorimotor areas and the superior temporal cortices. The factor 3-related parcels are located on the bilateral occipital cortices. The factor 4-related parcels are located on the bilateral lateral prefrontal cortices and supramarginal gyri. The factor 5-related parcels are located on the bilateral medial prefrontal cortices. The factor 6-related parcels are located on the bilateral precuneus areas and the inferior parietal cortices.

Diversity of Information Sources/Receivers Is Dependent on Cognitive Factors

Some cognitive functions may need various kinds of information to be realized while others may require only limited kinds of information. Similarly, information derived from some cognitive functions may be required to realize various kinds of cognitive functions while other information may be needed only from a small number of cognitive functions. We considered the diversity of informational interactions to be dependent on cognitive factors. Therefore, we quantified the diversity of information sources or receivers that were collected by parcels in the parcel network. To clarify our method, we assumptively describe some parcels, functions, or factors as *sources* in the present section. However, we note that these may be *receivers* because our method did not identify the directions of informational interactions.

First, for each parcel, we defined information sent from a cognitive factor as steady pseudo-temperature calculated from the heat diffusion equation in the network with heat sources whose intensities were defined by the NMF coefficient vector (column vector of the NMF coefficient matrix) (**Figure 6A**). This resulted in a temperature distribution over the parcel network for each cognitive factor. We observed that the temperatures of some parcels were roughly uniform across all cognitive functions, which implies equal collection of information. On the other hand, in another parcel, only one factor provided a high temperature and the other factors provided low temperatures, which implies a polarized collection of information. To quantify the degree of polarization, we used the Gini coefficients of the distributions of temperatures across cognitive factors (**Figure 6B**). Smaller Gini coefficients express more uniformity over cognitive factors, suggesting more diverse information sources/receivers.

Moreover, we investigated the 109 cognitive functions in terms of the diversity of information sources/receivers. For each cognitive function (or CFM), we averaged the Gini coefficients of the parcels whose parcel-to-CFM RSFCs were among the top ten (**Figure 6C**). The resulting value was regarded as the

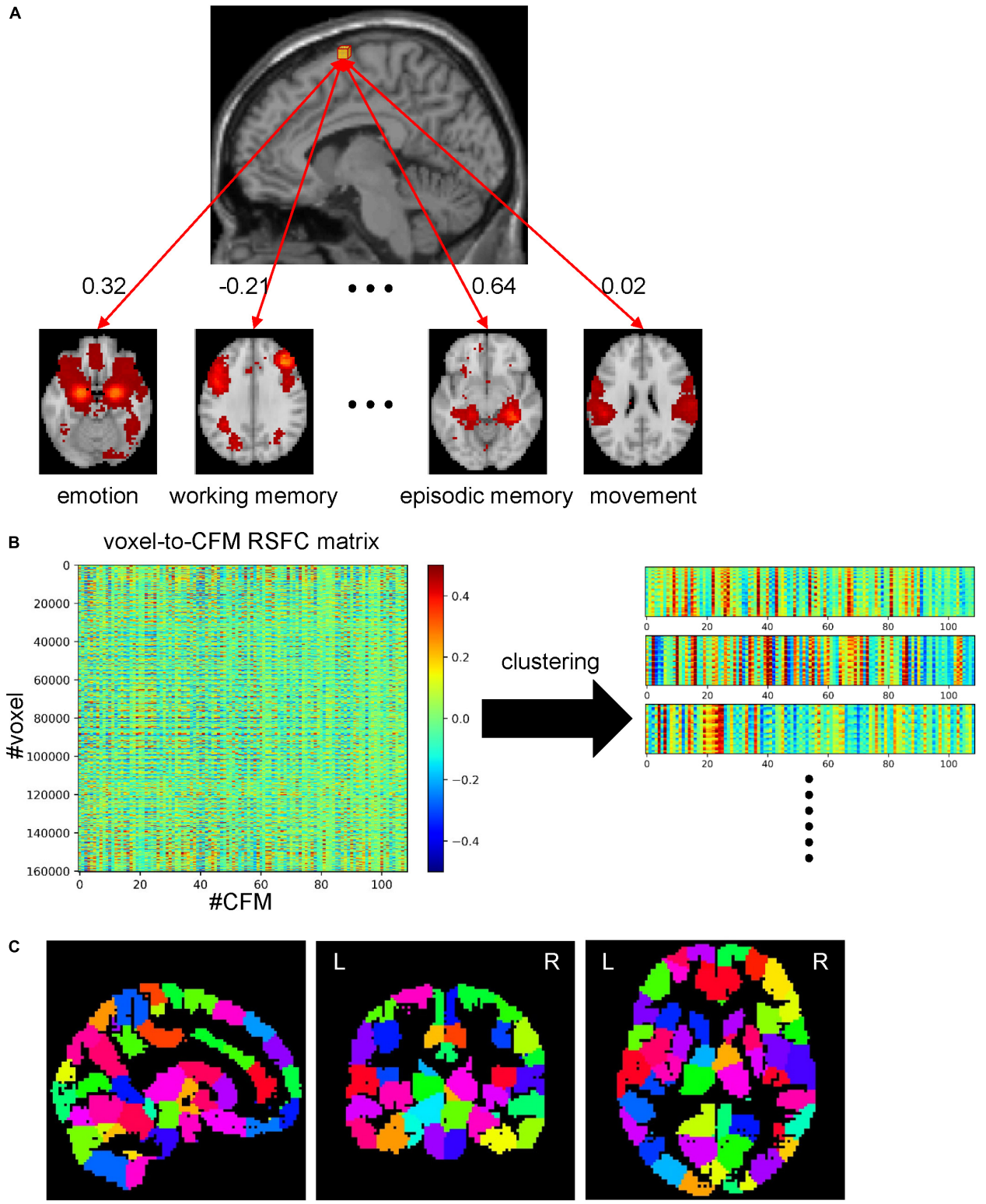


FIGURE 3 | Whole-brain parcellation based on cognitive function maps (CFMs). **(A)** Schematic illustration of voxel-to-CFM resting-state functional connectivities (RSFCs). A correlation coefficient between resting-state activities of each voxel and CFM was calculated, and was defined as the RSFC between them. **(B)** Parcellation was obtained by applying spectral clustering to the whole-brain voxel-to-CFM RSFC matrix. Each panel on the right corresponds to each resulting parcel. **(C)** The resulting parcellation consisting of 199 parcels.

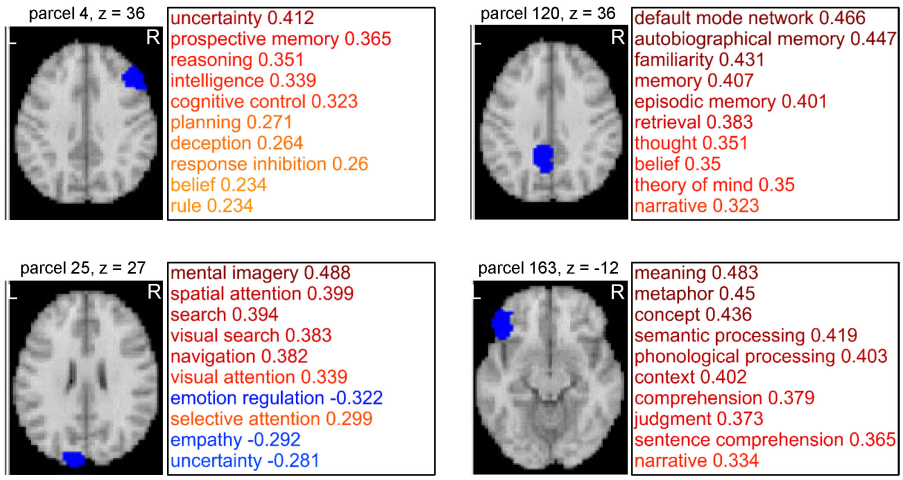


FIGURE 4 | Examples of parcels resulting from cognitive function-based whole-brain parcellation. Four examples of the parcels. Maps on a standard brain (left in each parcel) and top 10 relatedness with cognitive functions (right in each panel) are shown. The relatedness was defined as mean voxel-to-cognitive function map resting-state functional connectivities over the voxels belonging to the parcel.

Gini coefficient for the corresponding cognitive function. Upon sorting the NMF bases by the Gini coefficients of the cognitive functions, we observed cognitive factor-dependent differences in the diversity of information sources/receivers (**Figure 6D** and **Supplementary Table S9**). The factor 6-related cognitive functions tended to collect information from the most diverse sources/receivers. The factor 5- and 4-related functions had the second- and third-most diverse information sources/receivers, respectively. The diversity of information sources/receivers for the factor 1-related functions was moderate. The factor 2- and 3-related functions collected information from the most polarized sources/receivers.

The method used in this section has two parameters: diffusion coefficient D and decay constant β. Therefore, as an additional analysis, we investigated the effects of those parameter values. Since the result is only dependent on the ratio of those parameters, we only varied the diffusion coefficient D. As we observed, the Gini coefficients for the cognitive functions highly loaded by some factors were small and others were large. Thus, we compared the weighted sums of the Gini coefficients that express the means of the Gini coefficients for the cognitive functions assigned to the factors (see section Materials and Methods) between factors (**Figure 7**). Throughout the parameter region, we found qualitatively similar results to the one shown above except for the factor 3 that relates to vision and attention. The value of the weighted sum of the Gini coefficients for the factor 3 was largest when the diffusion coefficient was small, which means that the diversity of information sources/receivers was lowest. However, the diversity (indexed with the weighted sum of the Gini coefficients) relative to the others increased with an increase in the diffusion coefficient, and, finally became highest. Since the diffusion coefficient decides the range of information transmission, this result suggests the factor 3 (relating vision and attention) changes the relative diversity of informational interactions depending on the state of information transmission.

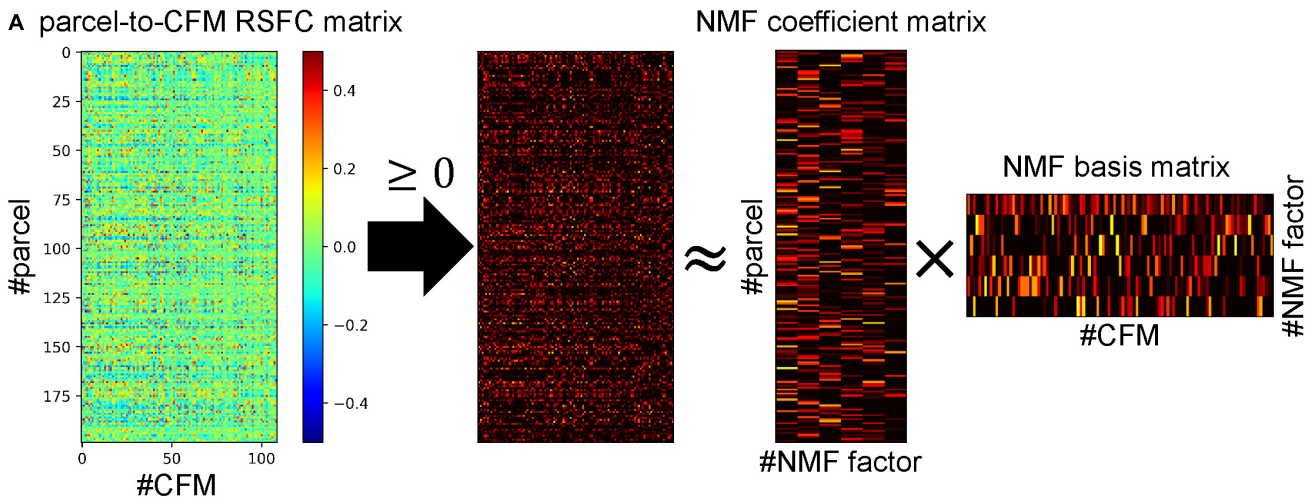
Cognitive Factor-Dependent Difference in Densities of Local Connectivity

The connection density of network which processes a cognitive function is an important factor to specify computational characteristics of the function. Using the clique percolation method, we identified local subnetworks within the parcels that were densely connected (**Figure 8**). By increasing the clique threshold K, subnetworks whose connectivity were denser came to the surface. When K was set to 8, we identified three densely connected subnetworks. By extracting the NMF coefficients for the parcels belonging to densely connected subnetworks, we found that these subnetworks were highly related with the factors 1 (blue), 2 (yellow), and 3 (green). The parcels composing each subnetwork are shown with the anatomical information in **Supplementary Table S10**.

The blue densely connected subnetwork includes the following regions: the left temporal cortex, left inferior parietal cortex, left supramarginal gyrus, left orbitofrontal cortex, left inferior frontal cortex (pars triangularis and pars orbitalis), left superior frontal cortex, left rostral middle frontal cortex, left anterior cingulate cortex, left frontal pole, and a small part of the left temporal pole.

The yellow densely connected subnetwork includes the following regions: the bilateral putamen, bilateral pallidum, bilateral caudal anterior cingulate cortices, bilateral posterior cingulate cortices, left middle temporal gyrus, bilateral superior temporal gyri, bilateral transverse temporal gyri, bilateral superior parietal cortices, bilateral supramarginal gyri, bilateral precuneus, bilateral precentral gyri, bilateral postcentral gyri, bilateral paracentral lobules, bilateral insula, bilateral pars opercularis (mainly left), and the bilateral superior frontal gyri (slightly lateralized to the left hemisphere).

The green densely connected subnetwork includes the following regions: the bilateral cerebellum, bilateral lateral



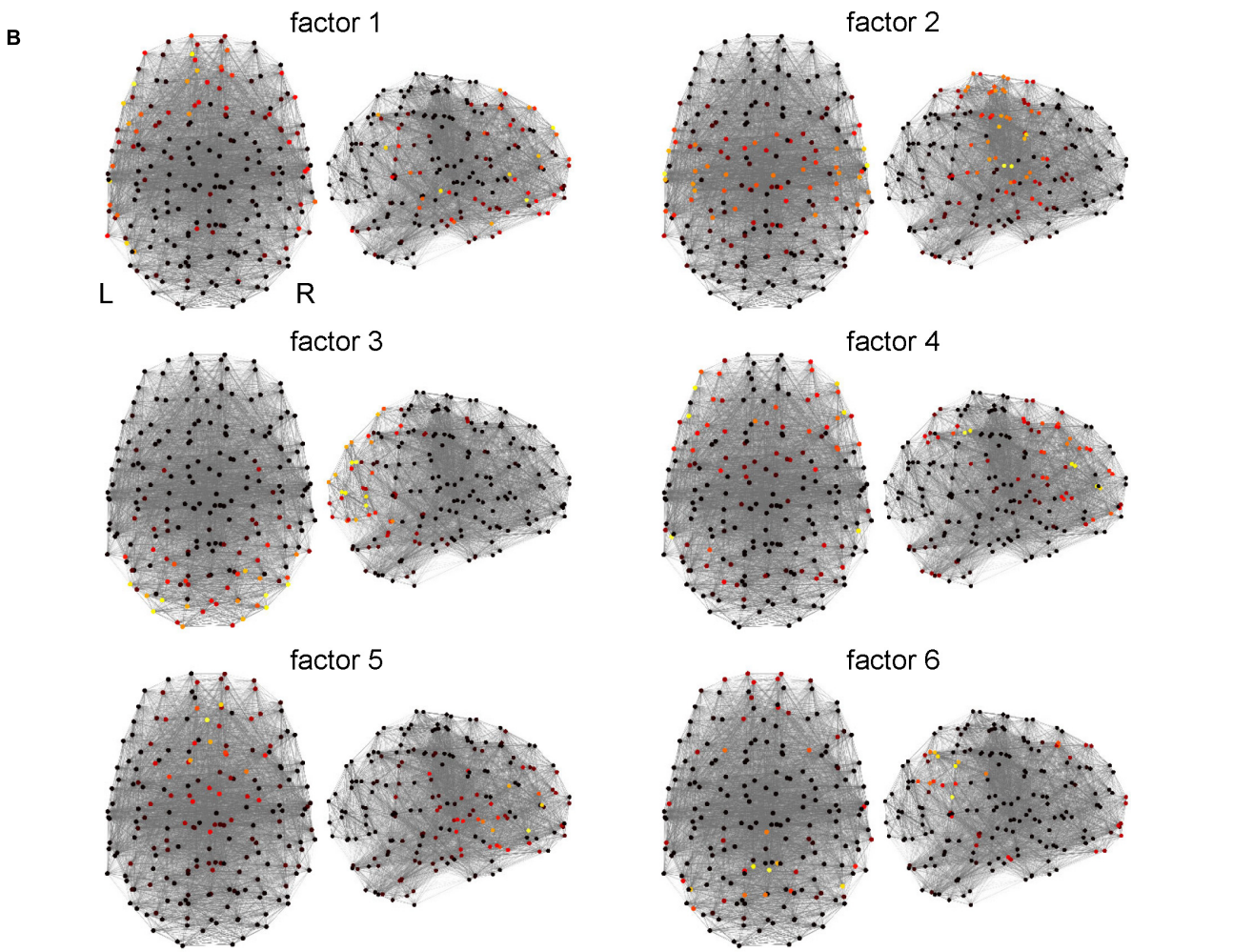


FIGURE 5 | Identifying cognitive factors using non-negative matrix factorization (NMF). **(A)** Procedure for identification of cognitive factors. First, the parcel-to-cognitive function map (CFM) resting-state functional connectivity matrix was thresholded with zero. Then, NMF was applied to the thresholded matrix, and the NMF coefficient and basis matrices were identified. The rows and columns of the NMF coefficient matrix correspond to parcels and NMF factors (cognitive factors), respectively. The rows and columns of the NMF basis matrix correspond to factors and CFMs, respectively. **(B)** The NMF coefficient values for each NMF factor are mapped on the parcels (shown as dots) located according to coordinates on a standard brain using heat mapping.

TABLE 2 | Cognitive factors defined using non-negative matrix factorization of the parcel-to-CFM RSFC matrix.

Factor 1 (concept processing)		Factor 2 (action and expression)		Factor 3 (vision and attention)	
Comprehension	0.248	Movement	0.329	Mental imagery	0.342
Narrative	0.245	Motor imagery	0.314	Spatial attention	0.333
Concept	0.244	Speech production	0.302	Visual search	0.329
Judgment	0.227	Skill	0.283	Search	0.317
Metaphor	0.221	Speech perception	0.256	Object recognition	0.252
Theory of mind	0.211	Motor control	0.245	Attention	0.242
Inference	0.204	Melody	0.235	Gaze	0.241
Belief	0.204	Integration	0.226	Face perception	0.223
Intention	0.202	Prosody	0.213	Selective attention	0.218
Semantic processing	0.191	Listening	0.207	Navigation	0.204
Factor 4 (executive function)		**Factor 5 (value and judgment)**		**Factor 6 (memory)**	
Cognitive control	0.303	Reward	0.320	Episodic memory	0.342
Rule	0.291	Anticipation	0.270	Default mode network	0.302
Working memory	0.289	Fear	0.263	Memory	0.293
Planning	0.288	Arousal	0.261	Autobiographical memory	0.278
Maintenance	0.276	Choice	0.255	Memory retrieval	0.266
Response inhibition	0.241	Decision making	0.233	Remembering	0.264
Expectancy	0.224	Loss	0.229	Retrieval	0.264
Task switching	0.216	Risk	0.225	Thought	0.262
Decision	0.210	Stress	0.224	Familiarity	0.254
Deception	0.198	Eating	0.202	Prospective memory	0.195

For each factor, the cognitive functions having the ten largest NMF basis values are shown with the corresponding NMF basis values. CFM, cognitive function map; RSFC, resting-state functional connectivity; NMF, non-negative matrix factorization.

occipital cortices, bilateral cuneus, bilateral pericalcarine cortices, bilateral lingual gyri, bilateral fusiform gyri, bilateral inferior parietal cortices, and the bilateral superior parietal cortices. Additionally, a small part of the inferior temporal cortex is included.

Network Communities That Are Uniformly or Diversely Associated With Cognitive Factors

Previous studies suggest that the RSFC network has a modular or community structure (He et al., 2009; Power et al., 2011; Bertolero et al., 2015, 2018). Such a community is considered as a module of information processing. To elucidate the information processing executed in each community, it is important to reveal whether the community is related to uniform or diverse kinds of cognitive functions. To this end, we identified the community structure by applying spectral clustering to the parcel-to-parcel RSFC matrix and investigated the functional uniformity or diversity of each community (**Figure 9**). The number of communities was set to 10, which maximized the silhouette coefficients. The NMF coefficients for the parcels belonging to the identified communities showed uniformity and diversity in their association with the cognitive factors in a community-dependent manner. The communities 2 and 8 specifically associated with cognitive factors 3 and 2, respectively. Conversely, the community 4, which was mainly located in the cerebellum, associated with diverse cognitive factors.

DISCUSSION

In the present study, we endeavored to show a whole picture of the human cognition and to reveal characteristics of each cognitive function that constitutes it. To this end, we investigated the relationships among 109 cognitive functions based on two ideas: (1) the cognitive functions that overlapping brain regions are responsible for should be interrelated, and (2) the cognitive functions that connected brain regions are responsible for should be also interrelated. Especially, we characterized 109 cognitive functions based on the CFM and RSFC-determined relationships among them. First, we presented a relational mapping that involved two-dimensional embedding of the cognitive functions using the RSFCs among CFMs. Then, we performed conceptual analysis in which a cognitive function was analyzed to identify the subfunctions constituting it, based on the RSFCs between voxels in the targeted CFM and the remaining CFMs. Moreover, we obtained a novel whole-brain parcellation in which each parcel had the vector of relatedness with these cognitive functions. Based on the network analyses using the parcels, we identified six cognitive factors, quantified the diversity of information sources/receivers for each cognitive function and factor, found the densely connected subnetworks associated with specific cognitive factors, and identified the communities that were associated with uniform or diverse cognitive factors. Altogether, we suggest the effectiveness of our approach in which we combined a large-scale meta-analysis of functional brain mapping with the methods of network neuroscience

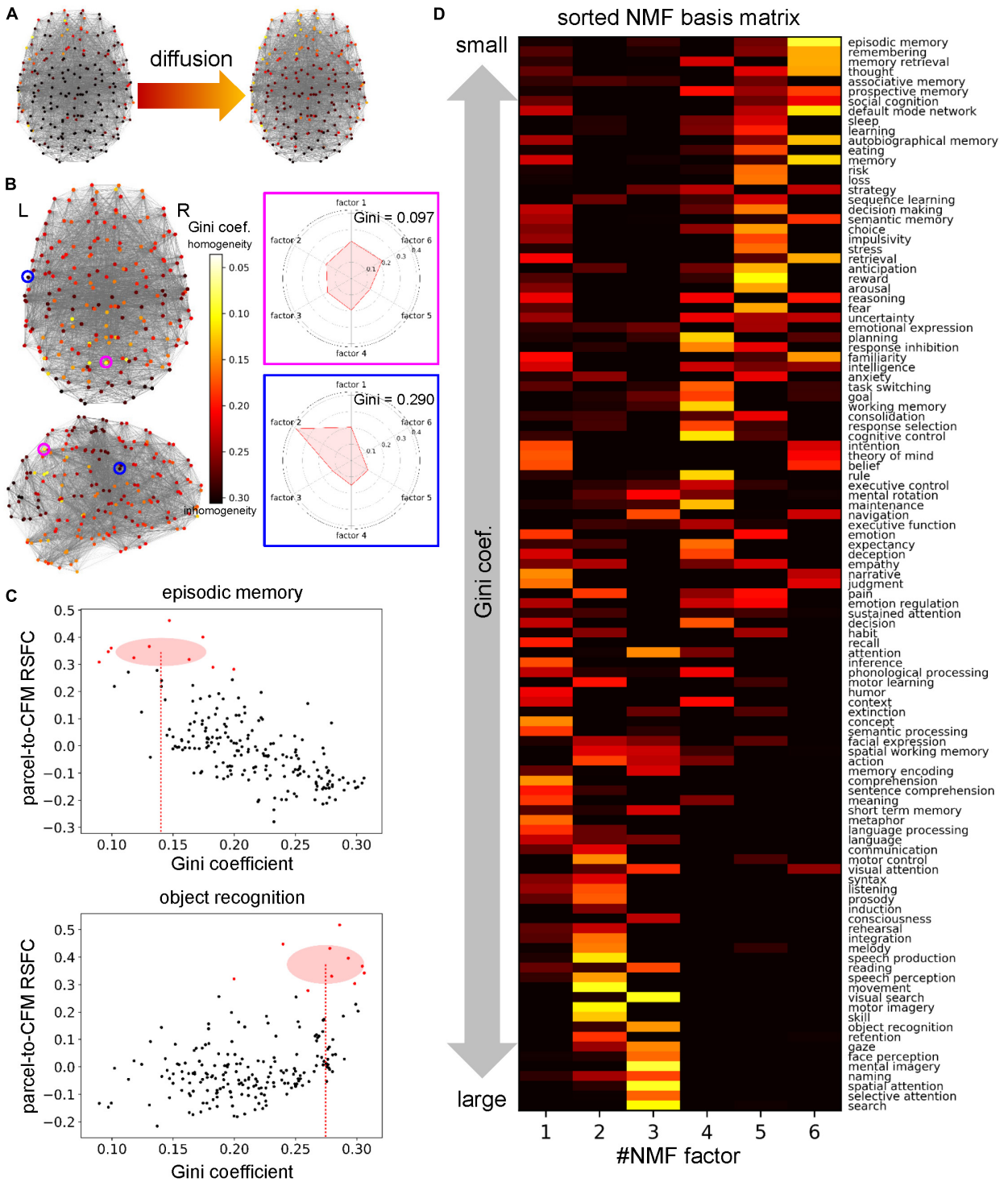


FIGURE 6 | Cognitive factor-dependent diversity of informational interactions. **(A)** Heat source (left) and temperature (right) distributions on the parcels for the cognitive factor 1. The heat source values were defined as the non-negative matrix factorization (NMF) coefficient values of the corresponding column. Temperatures were calculated at steady states of the diffusion process governed by the graph Laplacian. **(B)** The Gini coefficient distribution. For each parcel, the Gini coefficient represents inhomogeneity of temperatures across the factors. The Gini coefficients and polar graph of temperature for the parcel circled with magenta and blue are shown on the right. **(C)** Example plots between the Gini coefficient values and parcel-to-cognitive function map (CFM) resting-state functional connectivities (RSFCs). Upper and lower plots correspond to the CFMs of 'episodic memory' and 'object recognition,' respectively. Each dot expresses each parcel. The parcels that have the 10 largest RSFCs are red-colored. The means and standard deviations of the Gini coefficients for these red-colored parcels are shown as the centers and radiuses of the red circles, respectively. The means are also indicated by the red dotted line. **(D)** The transposed NMF basis matrix, sorted by the Gini coefficients. The cognitive function corresponding to each CFM is shown on the right.

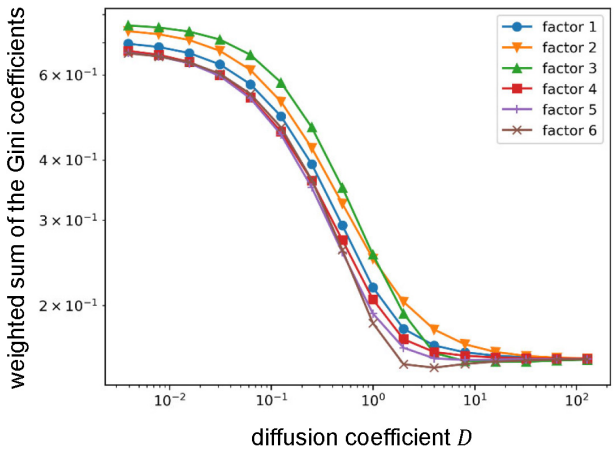


FIGURE 7 | Effect of the value of the diffusion coefficient on the weighted sum of the Gini coefficient. Log-log plots of the weighted sums of Gini coefficients with varying the diffusion coefficient. Throughout the parameter region, relative orders are qualitatively similar except for the factor 3. The weighted sum of Gini coefficients for factor 3 moves from largest (i.e., lowest diversity of informational interaction) to smallest (i.e., highest diversity of informational interaction).

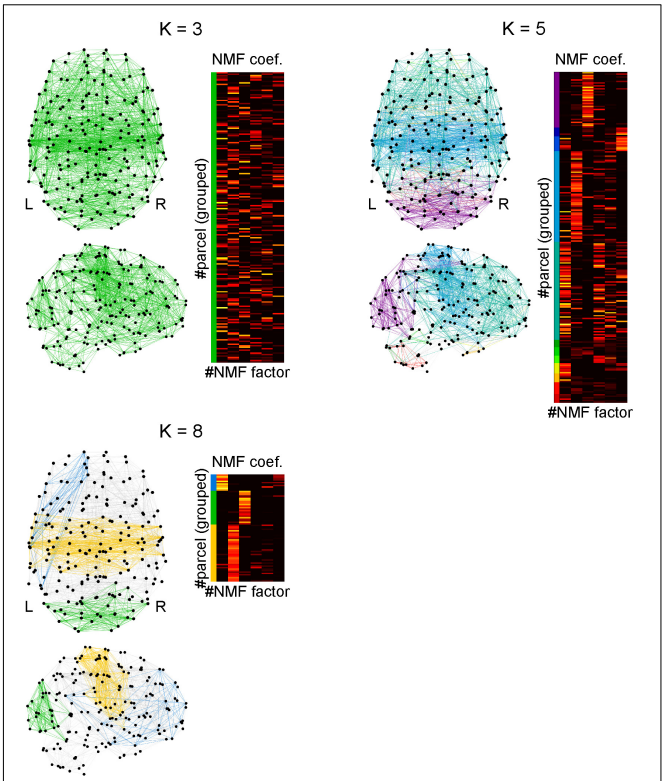


FIGURE 8 | Identifying densely connected subnetworks using clique percolation method. Subnetworks are color-coded and the non-negative matrix factorization coefficients of the parcels belonging to the subnetworks are shown in corresponding colors. The clique threshold K is a criterion for densities of connectivity in subnetworks to be identified. The connectivity becomes dense with an increase in K. When $K = 3$, all parcels were interconnected, which implies only one network was identified. When $K = 5$, 11 subnetworks were identified. When $K = 8$, three subnetworks were identified, in each of which the parcels were densely interconnected.

to investigate the relationships among cognitive functions to understand each cognitive function *per se* and the human as a relational system consisting of cognitive functions.

Implications of the Results and Comparisons With Previous Studies

Categorization of cognitive functions is an essential first step not only for the scientific understanding of the brain but also for the clinical application of neuroscientific knowledge for diagnosis of psychiatric diseases. In this study, we provided such categorizations using two methods. One was based on the clustering on the CFM-to-CFM network and also yielded six cognitive clusters, including 'language,' 'action and expression,' 'vision and attention,' 'executive function,' 'value and judgment,' and 'self and others.' The other was based on the NMF, and yielded six cognitive factors: 'concept processing,' 'action and expression,' 'vision and attention,' 'executive function,' 'value and judgment,' and 'memory.'

We show the entire correspondences between the cognitive clusters and the factors in **Figure 10**. The cognitive factors 'action and expression,' 'vision and attention,' 'executive function,' and 'value and judgment' roughly correspond to the cognitive clusters that are labeled with the same names. The 'memory' factor mainly relates to the 'self and others' cluster. Additionally, we like to stress that several functions that are strongly associated with the 'memory' factor (e.g., 'memory retrieval' and 'prospective memory') belong to the 'executive function' cluster. The 'concept processing' factor seems to relate to the cognitive clusters in a complex manner. Considering the NMF basis vector, it is suggested to be related to both 'language' and 'self and others' clusters. Therefore, the concepts of 'memory,' 'concept processing,' 'executive function,' 'language,' and 'self and others' are entangled, and the information processing relating these concepts may be executed through close interactions among them.

In the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), which describes the current standardized criteria to diagnose psychiatric diseases, the neurocognitive domain is categorized into six subdomains consisting of 'complex attention,' 'executive function,' 'learning and memory,' 'language,' 'perceptual-motor,' and 'social cognition' (American Psychiatric Association, 2013). We found rough correspondences between the categorizations in DSM-5 and our results. The 'complex attention' subdomain in DSM-5 is considered to be included in the 'vision and attention' cognitive factor and cluster in the present study. The 'executive function' subdomain in DSM-5 probably corresponds to the cognitive factor and cluster labeled with the same name in this study. The 'learning and memory' subdomain in DSM-5 mainly relates to the 'memory' factor in this study. Since the immediate memory is included in the 'learning and memory' subdomain in DSM-5, this may relate to the 'executive function' cognitive factor and cluster in this

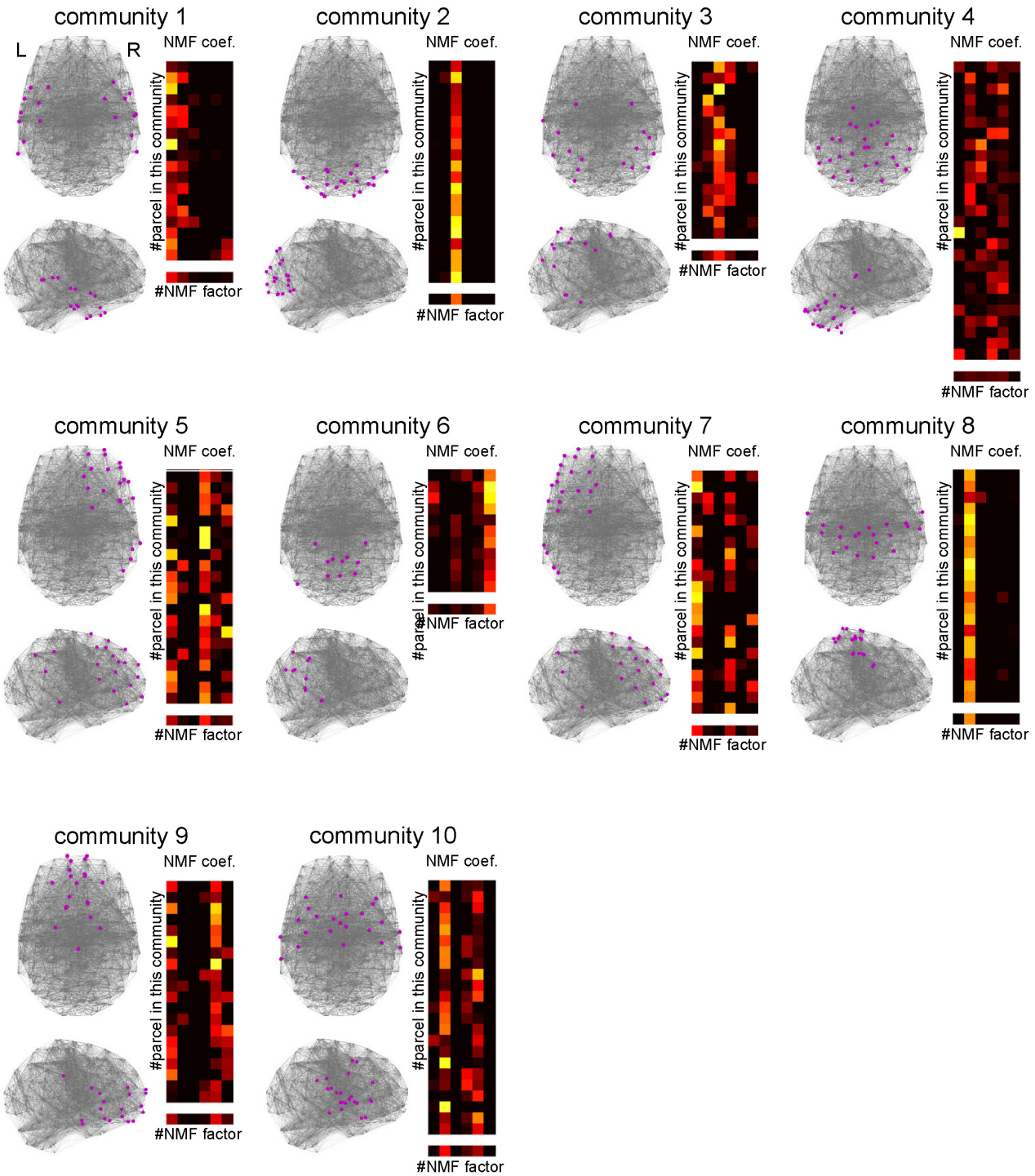


FIGURE 9 | Characterizing communities in the whole-brain network with the cognitive factors. Ten network communities were identified using spectral clustering. The parcels belonging to each community are shown by magenta dots. For each community, the subpart of the non-negative matrix factorization coefficient matrix corresponding to the parcels in the community is shown on the right. The lowest row shows the row mean of the matrix.

study that involves 'maintenance' and 'working memory.' The 'language' subdomain in DSM-5 roughly corresponds to the 'language' cluster in our analysis. Furthermore, it also relates to the 'action and expression' cluster in this study because it includes 'syntax,' 'listening,' 'communication,' and so on. Additionally, the 'language' subdomain in DSM-5 probably has a close relationship with the 'concept processing' and 'action and expression' factors in this study. The 'perceptual-motor' subdomain in DSM-5

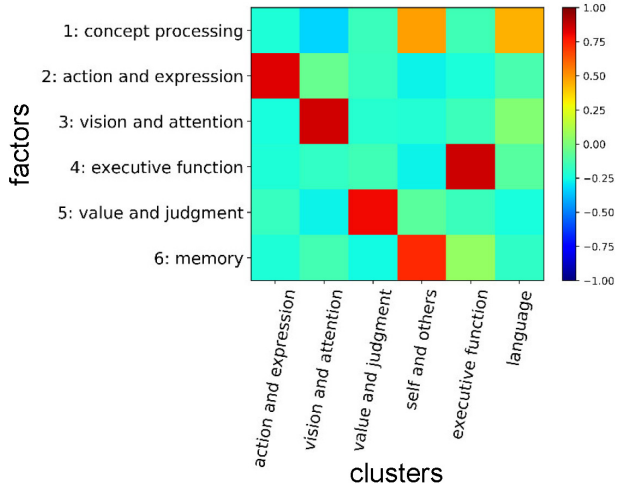


FIGURE 10 | Correspondences between cognitive clusters and factors. Correlation coefficients between NMF basis vectors for the factors and presence/absence vectors of cognitive functions for the clusters are shown. In the presence/absence vectors, the presences and absences in the cluster were assigned to '1' and '0', respectively.

mainly relates to the 'action and expression' and 'vision and attention' factors and clusters in this study. The 'social cognition' subdomain in DSM-5 mainly relates to the 'value and judgment' and 'self and others' clusters in this study. It may also relate to the 'concept processing' and 'value and judgment' factors.

The relational mapping among cognitive functions that we obtained provides several insights into the mechanisms of cognition. We found that the default-mode network was located in a position close to 'self and others'-related cognitive functions (e.g., 'theory of mind' and 'autobiographical memory') and social cognitive functions (e.g., 'social cognition' and 'decision-making'). In fact, many studies suggest that these cognitive functions share underlying neural substrates (Spreng et al., 2009; Andrews-Hanna et al., 2010, 2014; Spreng and Grady, 2010; Mars et al., 2012; Reniers et al., 2012; Li et al., 2014; Meyer et al., 2019). We also found that 'phonological processing' was located close to the 'executive function' cluster. This seems to be consistent with Baddeley's working memory system (Baddeley, 2000), in which phonological loop interacts with central execution. From the same point of view, we can link 'episodic memory' with episodic buffer in Baddeley's system, since it was also located close to the 'executive function' cluster. More globally, we observed that the 'executive function' cluster neighbored the 'self and others' cluster, centering on the 'default-mode network.' Several studies reported cooperative activity between the brain areas related to these cognitive functions when subjects experienced spontaneous thoughts (Christoff et al., 2009) and engaged in creative tasks (Beaty et al., 2015) and mental simulations (Gerlach et al., 2011). Thus, our relational mapping of cognitive functions provides a whole picture of cognition which is feasible because it includes many known neurocognitive relationships. A study to survey relationships among cognitive functions whose aim was similar to ours was conducted using text analysis of neuroscience literature (Beam et al., 2014). In this study, the authors identified networks among 100 cognitive concepts, among 100 anatomical regions, and among combinations of both on the basis of the co-occurrences of the terms in the texts. More recently, a study reported the relations among 120 cognitive functions using hierarchical clustering based on correlations between pseudo-activation patterns, not RSFCs (Alexander-Bloch et al., 2018). Owing to methodological variations between the present and those studies, the present study can endow another picture complementing these studies.

In the present study, we proposed a novel method for conceptual analysis of cognitive concepts based on the CFMs and RSFCs in the brain. This yielded functional subdivisions of the cognitive concepts. Each sub-concept was characterized by its relatedness with the other cognitive concepts. We found several unexpectedly characterized sub-concepts. A sub-concept of 'emotion' that is characterized by functionality involving comprehension of abstract meanings is one such unexpected sub-concept. This may imply that we need emotional processing to receive an implicit message from linguistic expressions. Conversely, emotional processing may require analysis of abstract meanings. Further, we found that 'thought' had a sub-concept related to imaginary navigation. Navigation is considered to be handled by the grid and place cell systems. Several studies have shown that these systems play roles not only in physical spaces but also in abstract spaces such as social relationships, features of objects and events, and relational knowledge (Tavares et al., 2015; Constantinescu et al., 2016; Epstein et al., 2017; Garvert et al., 2017; Aronov et al., 2017; Schafer and Schiller, 2018). Therefore, imaginary navigation in an abstract space may be generally used in thoughts.

In the analysis for diversity of informational interactions, we observed that the nodes associated with cognitive functions that were closely related to the 'memory' factor interacted with the most diverse information. Since our analyses did not indicate the directions of the interactions, it was not clear whether these nodes were information sources or receivers. If the nodes play the role of information source, our result suggests that information processed with 'memory'-related functions is necessary to realize a wide range of cognitive functions. Conversely, if the nodes are receivers of information, it suggests that execution of 'memory'-related cognitive functions need information from a wide range of cognitive functions. Since the 'value and judgment'- and 'executive function'-related cognitive functions also have relatively diverse informational interactions, these results suggest similar implications. To support these results, an analysis to clarify the interaction directions will be required. In the additional analysis, the behavior of the factor relating to 'vision and attention' is insightful since it suggests that the diversity of informational interaction highly depends on the range of information transmission. Since the efficacy of information transmission changes depending on the brain state such as wakefulness and sleep (Massimini et al., 2005), our observation may suggest that the role of visual and attentional processing on the entire cognitive information processing changes when the brain state shifts.

We identified a densely connected subnetwork that was highly related to the 'concept processing' factor as well as the subnetworks related to the 'action and expression' and 'vision and attention' factors. The 'concept processing' subnetwork included a direct pathway between the Broca's and Wernicke's areas and an indirect pathway passing through the left inferior parietal cortex, which has been previously identified as constituents of the perisylvian language networks (Catani et al., 2005). Moreover, we detected participation of a wide range of structures in the left prefrontal cortex, including the lateral, medial, and orbital regions as well as the frontal pole in this subnetwork. Since these areas involve various aspects of higher-order cognition (Passingham and Wise, 2012; Fuster, 2015), this subnetwork suggests the existence of an integrated cognitive function that is highly dependent on language processing but is contributed also from functions beyond language processing.

Previous studies have shown that functional communities exist in the brain (He et al., 2009; Eickhoff et al., 2011; Power et al., 2011; Crossley et al., 2013; Bertolero et al., 2015, 2018). The studies have emphasized functional specificities of the communities. On the other hand, we found differences in the degrees of functional specificities of the communities, in which some communities were specifically associated with one cognitive factor while other communities were associated with diverse cognitive factors. One of the communities associated with the most diverse cognitive factors was located mainly in the cerebellum. Although the cerebellum was previously considered to be related to motor functions, it is now recognized that the cerebellum involves a remarkably wide range of cognitive functions (Stoodley and Schmahmann, 2009; Strick et al., 2009; Stoodley, 2012), which is consistent with our results. Viewing the internal models in the cerebellum (Wolpert et al., 1998) as a general controller working on various mental activities may give rise to a theoretical foundation for the diversity of cerebellar functionality (Ito, 2008). Additionally, a theoretical study (Yamazaki and Tanaka, 2007) suggests that the cerebellum is considered a kind of universal machine, the so-called liquid state machine (Maass et al., 2002), which may also support our finding.

Limitations and Future Directions

There are several limitations to the present study which should be addressed in future studies. While constructing the CFMs, we used abstract texts to count the occurrences of cognitive terms. We did not utilize contextual information. Therefore, we did not discriminate as to whether the occurrences meant activation or deactivation. Additionally, to ensure that a term was the main topic in a study, we only used the frequency of the occurrences in its title, abstract, and keywords. Utilization of contextual information is a promising way to improve our analyses. The methods being developed in the field of natural language processing will probably provide such ways. Additionally, the use of natural language processing technics can provide us useful data revealing the constraints of inferring relationships among cognitive functions.

Compared to the datasets stored in the recently developing large-scale databases such as the Human Connectome Project database (Smith et al., 2013; Van Essen et al., 2013), the dataset used in the present study was small with respect to both the number of subjects and the number of scan volumes. Although we checked the reliabilities of the RSFC matrices and we consider that the outlines of the results are validated, especially in details of the results, some instabilities caused by the small data size were probably not removed. Therefore, we should continuously revise and establish knowledge suggested from our observations.

The number of parcels in the cognitive function-based whole-brain parcellation was determined not based on data but by reference to several existing atlases (Destrieux et al., 2010; Power et al., 2011; Shen et al., 2013; Baldassano et al., 2015; Fan et al., 2016). The selection of the number is a trade-off problem. The larger number of parcels results in a set of smaller parcels. This is suitable to reflect spatial heterogeneity in the brain. On the other hand, since the BOLD signal of the parcel is calculated by averaging the signals over the voxels within it, the signal of a smaller parcel tends to be more fluctuated. Several studies suggest that estimations of network characteristics in the brain depend on a resolution of parcellation (de Reus and van den Heuvel, 2013; Proix et al., 2016). Therefore, we need to address the issue of the number of parcels in the future study.

The connectivity measures used in this study were undirected and did not provide any information regarding dynamic causality and logical orders. On the other hand, we expect that the identification of directions in connectivity will provide useful insights into the issues dealt with in the present study. One representative instance is the analysis of the diversity of informational interactions, in which we found that cognitive functions highly related to the 'memory' factor interact with the most diverse kinds of information. If those cognitive functions are information sources (or receivers), this result suggests certain roles (or mechanisms) of memory-related information processing in the entire cognitive information processing. Similarly, directional information is required in the RSFC-based conceptual analysis, in which we observed that 'emotion' should implicate a sub-concept related to comprehension of abstract meanings. To determine whether emotional processing contributes to comprehension or vice versa, we need to identify the direction of connectivity between the CFM corresponding to 'emotion' and the CFMs of the cognitive functions related to comprehension of abstract meanings. Another instance in which directional information is required is the relational mapping of cognitive functions. This is expected to reveal the hierarchical dependencies among the cognitive functions, which will provide a more sophisticated perspective for the mechanism of the entire human cognition. To these ends, we may use methods of time series analyses, including the dynamic causal modeling (Friston et al., 2003, 2014), Granger causality (Roebroeck et al., 2005; Seth, 2010), and transfer entropy (Schreiber, 2000; Vicente et al., 2011).

In our relational mapping, we used multidimensional scaling to embed cognitive functions. Although this method provided

an easily interpretable overview of the relationship among cognitive functions, the distances between them were more or less distorted. Therefore, we need more sophisticated embedding methods. The t-SNE may be such a method (van der Maaten and Hinton, 2008; van der Maaten, 2014). Recently, embedding methods into non-Euclidean spaces, such as Poincaré embedding, has been proposed (Nickel and Kiela, 2017). Such a non-Euclidean embedding method is considered to reveal other types of information regarding the relationships among cognitive functions. In addition, on the basis of CFMs, RSFCs, and other useful neuroscientific tools, exploring ontological relations [e.g., *is-a* and *part-of* relationships (Lenartowicz et al., 2010; Hastings et al., 2014; Poldrack and Yarkoni, 2016)] is an important future direction.

The methods and results provided in the present study let us clarify the meaning of each cognitive concept and obtain an analytic and synthetic understanding of the relationships among cognitive concepts. This possibly provides an empirical sketch of the research domains of cognitive neuroscience, which has been the aim of neuroimaging studies involving meta-analytical methods (Alhazmi et al., 2018). Moreover, this will stimulate the research fields of biological brain- and/or cognition-inspired artificial intelligences (Anderson and Lebiereeds, 1998; Anderson et al., 2004; Anderson, 2005; Eliasmith et al., 2012; Hassabis et al., 2017) by providing guidelines for understanding human cognition as a whole.

DATA AVAILABILITY STATEMENT

Raw data cannot be shared publicly because the sharing of raw data was not ethically approved. Data without any personally identifiable information will be available from author HK's web page.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the institutional ethics committee of the National Center of Neurology and Psychiatry (NCNP). The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HK performed the experiments and data analysis. All authors contributed to the design of the study and writing of the manuscript.

FUNDING

This research was partially supported by a JSPS Grant-in-Aid for Young Scientists (B) (26870934), the Sasakawa Scientific Research Grant from The Japan Science Society, and the Leading Initiative for Excellent Young Researchers, MEXT, Japan.

ACKNOWLEDGMENTS

We would like to thank Editage (www.editage.jp) for English language editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2019.00457/full#supplementary-material

REFERENCES

Alexander-Bloch, A., Shou, H., Liu, S., Satterthwaite, T. D., Glahn, D. C., Shinohara, R. T., et al. (2018). On testing for spatial correspondence between maps of human brain structure and function. *Neuroimage* 178, 540–551. doi: 10.1016/j.neuroimage.2018.05.070

Alhazmi, F. H., Beaton, D., and Abdi, H. (2018). Semantically defined subdomains of functional neuroimaging literature and their corresponding brain regions. *Hum. Brain Mapp.* 39, 2764–2776. doi: 10.1002/hbm.24038

American Psychiatric Association, (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*, 5th Edn, Washington, DC: American Psychiatric Association Pub.

Anderson, J. R. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cogn. Sci.* 29, 313–341. doi: 10.1207/s15516709cog0000_22

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. (2004). An integrated theory of the mind. *Psychol. Rev.* 111, 1036–1060. doi: 10.1037/0033-295X.111.4.1036

Anderson, J. R., and Lebiere, C. J. eds (1998). *The Atomic Components of Thought*. Mahwah, NJ:Lawrence Erlbaum Associates.

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., and Buckner, R. L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron* 65, 550–562. doi: 10.1016/j.neuron.2010.02.005

Andrews-Hanna, J. R., Saxe, R., and Yarkoni, T. (2014). Contributions of episodic retrieval and mentalizing to autobiographical thought: evidence from functional neuroimaging, resting-state connectivity, and fMRI meta-analyses. *Neuroimage* 91, 324–335. doi: 10.1016/j.neuroimage.2014.01.032

Aronov, D., Nevers, R., and Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit. *Nature* 543, 719–722. doi: 10.1038/nature21692

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* 4, 417–423. doi: 10.1016/S1364-6613(00)01538-2

Baldassano, C., Beck, D. M., and Fei-Fei, L. (2015). Parcellating connectivity in spatial maps. *PeerJ* 3:e784. doi: 10.7717/peerj.784

Bassett, D. S., Bullmore, E., Verchinski, B. A., Mattay, V. S., Weinberger, D. R., and Meyer-Lindenberg, A. (2008). Hierarchical organization of human cortical networks in health and schizophrenia. *J. Neurosci.* 28, 9239–9248. doi: 10.1523/JNEUROSCI.1929-08.2008

Bassett, D. S., and Sporns, O. (2017). Network neuroscience. *Nat. Neurosci.* 20, 353–364. doi: 10.1038/nn.4502

Beam, E., Appelbaum, L. G., Jack, J., Moody, J., and Huettel, S. A. (2014). Mapping the semantic structure of cognitive neuroscience. *J. Cogn. Neurosci.* 26, 1949–1965. doi: 10.1162/jocn_a_00604

Beaty, R. E., Benedek, M., Barry Kaufman, S., and Silvia, P. J. (2015). Default and executive network coupling supports creative idea production. *Sci. Rep.* 5:10964. doi: 10.1038/srep10964

Beaty, R. E., Silvia, P. J., Nusbaum, E. C., Jauk, E., and Benedek, M. (2014). The roles of associative and executive processes in creative cognition. *Mem. Cognit.* 42, 1186–1197. doi: 10.3758/s13421-014-0428-8

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300. doi: 10.2307/2346101

Bertolero, M. A., Yeo, B. T. T., Bassett, D. S., and D'Esposito, M. (2018). A mechanistic model of connector hubs, modularity and cognition. *Nat. Hum. Behav.* 2, 765–777. doi: 10.1038/s41562-018-0420-6

Bertolero, M. A., Yeo, B. T. T., and D'Esposito, M. (2015). The modular and integrative functional architecture of the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 112, E6798–E6807. doi: 10.1073/pnas.1510619112

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magn. Reson. Med.* 34, 537–541. doi: 10.1002/mrm.1910340409

Borg, I., and Groenen, P. J. F. (1997). *Modern Multidimensional Scaling: Theory and Applications*. New York, NY: Springer.

Boutsidis, C., and Gallopoulos, E. (2008). SVD based initialization: a head start for nonnegative matrix factorization. *Pattern Recognit.* 41, 1350–1362. doi: 10.1016/j.patcog.2007.09.010

Brandl, F., Le Houcq Corbi, Z., Bratec, S. M., and Sorg, C. (2019). Cognitive reward control recruits medial and lateral frontal cortices, which are also involved in cognitive emotion regulation a coordinate-based meta-analysis of fMRI studies. *Neuroimage* 200, 659–673. doi: 10.1016/j.neuroimage.2019.07.008

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198. doi: 10.1038/nrn2575

Bullmore, E., and Sporns, O. (2012). The economy of brain network organization. *Nat. Rev. Neurosci.* 13, 336–349. doi: 10.1038/nrn3214

Bzdok, D., Langner, R., Schilbach, L., Jakobs, O., Roski, C., Caspers, S., et al. (2013). Characterization of the temporo-parietal junction by combining data-driven parcellation, complementary connectivity analyses, and functional decoding. *Neuroimage* 81, 381–392. doi: 10.1016/j.neuroimage.2013.05.046

Catani, M., Jones, D. K., and Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Ann. Neurol.* 57, 8–16. doi: 10.1002/ana.20319

Chang, L. J., Yarkoni, T., Khaw, M. W., and Sanfey, A. G. (2013). Decoding the role of the insula in human cognition: functional parcellation and large-scale reverse inference. *Cereb. Cortex* 23, 739–749. doi: 10.1093/cercor/bhs065

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., and Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8719–8724. doi: 10.1073/pnas.0900234106

Chuderski, A., and Jastrzêbski, J. (2018). Much ado about aha!: insight problem solving is strongly related to working memory capacity and reasoning ability. *J. Exp. Psychol. Gen.* 147, 257–281. doi: 10.1037/xge0000378

Chung, F. (1997). *Spectral Graph Theory*. Providence, RI:American Mathematical Society.

Cieslik, E. C., Zilles, K., Caspers, S., Roski, C., Kellermann, T. S., Jakobs, O., et al. (2013). Is there "one" DLPFC in cognitive action control? evidence for heterogeneity from co-activation-based parcellation. *Cereb. Cortex* 23, 2677–2689. doi: 10.1093/cercor/bhs256

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. doi: 10.1093/bioinformatics/btp163

Constantinescu, A. O., OReilly, J. X., and Behrens, T. E. J. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science* 352, 1464–1468. doi: 10.1126/science.aaf0941

Cordes, D., Haughton, V. M., Arfanakis, K., Carew, J. D., Turski, P. A., Moritz, C. H., et al. (2001). Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data. *Am. J. Neuroradiol.* 22, 1326–1333

Craddock, R. C., James, G. A., Holtzheimer, P. E., Hu, X. P., and Mayberg, H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum. Brain Mapp.* 33, 1914–1928. doi: 10.1002/hbm.21333

Crossley, N. A., Mechelli, A., Vertes, P. E., Winton-Brown, T. T., Patel, A. X., Ginestet, C. E., et al. (2013). Cognitive relevance of the community structure of the human brain functional coactivation network. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11583–11588. doi: 10.1073/pnas.1220826110

de la Vega, A., Chang, L. J., Banich, M. T., Wager, T. D., and Yarkoni, T. (2016). Large-scale meta-analysis of human medial frontal cortex reveals tripartite functional organization. *J. Neurosci.* 36, 6553–6562. doi: 10.1523/JNEUROSCI.4402-15.2016

de Reus, M. A., and van den Heuvel, M. P. (2013). The parcellation-based connectome: limitations and extensions. *Neuroimage* 80, 397–404. doi: 10.1016/j.neuroimage.2013.03.053

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021

Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* 53, 1–15. doi: 10.1016/j.neuroimage.2010.06.010

Eickhoff, S. B., Bzdok, D., Laird, A. R., Roski, C., Caspers, S., Zilles, K., et al. (2011). Co-activation patterns distinguish cortical modules, their connectivity and functional differentiation. *Neuroimage* 57, 938–949. doi: 10.1016/j.neuroimage.2011.05.021

Eisenberg, I. W., Bissett, P. G., Zeynep Enkavi, A., Li, J., MacKinnon, D. P., Marsch, L. A., et al. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nat. Commun.* 10:2319. doi: 10.1038/s41467-019-10301-1

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., et al. (2012). A large-scale model of the functioning brain. *Science* 338, 1202–1205. doi: 10.1126/science.1225266

Epstein, R. A., Patai, E. Z., Julian, J. B., and Spiers, H. J. (2017). The cognitive map in humans: spatial navigation and beyond. *Nat. Neurosci.* 20, 1504–1513. doi: 10.1038/nn.4656

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., et al. (2016). The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cereb. Cortex* 26, 3508–3526. doi: 10.1093/cercor/bhw157

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355. doi: 10.1016/S0896-6273(02)00569-X

Fornito, A., Zalesky, A., and Bullmore, E. T. (2016). *Fundamentals of Brain Network Analysis*. London: Academic Press.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9673–9678. doi: 10.1073/pnas.0504136102

Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *Neuroimage* 19, 1273–1302. doi: 10.1016/S1053-8119(03)00202-7

Friston, K. J., Kahan, J., Biswal, B., and Razi, A. (2014). A DCM for resting state fMRI. *Neuroimage* 94, 396–407. doi: 10.1016/j.neuroimage.2013.12.009

Fuhrmann, D., Simpson-Kent, I. L., Bathelt, J., Team, T. C., and Kievit, R. A. (2019). A hierarchical watershed model of fluid intelligence in childhood and adolescence. *Cereb. Cortex* doi: 10.1093/cercor/bhz091 [Epub ahead of print].

Fuster, J. M. (2015). *The Prefrontal Cortex*, 5th Edn, London: Academic Press.

Garvert, M. M., Dolan, R. J., and Behrens, T. E. (2017). A map of abstract relational knowledge in the human hippocampal-entorhinal cortex. *elife* 6:e17086. doi: 10.7554/eLife.17086

Gerlach, K. D., Spreng, R. N., Gilmore, A. W., and Schacter, D. L. (2011). Solving future problems: default network and executive activity associated with goal-directed mental simulations. *Neuroimage* 55, 1816–1824. doi: 10.1016/j.neuroimage.2011.01.030

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature* 536, 171–178. doi: 10.1038/nature18933

Greve, D. N., and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* 48, 63–72. doi: 10.1016/j.neuroimage.2009.06.060

Guo, C. C., Kurth, F., Zhou, J., Mayer, E. A., Eickhoff, S. B., Kramer, J. H., et al. (2012). One-year test-retest reliability of intrinsic connectivity network fMRI in older adults. *Neuroimage* 61, 1471–1483. doi: 10.1016/j.neuroimage.2012.03.027

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., et al. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol.* 6:e159. doi: 10.1371/journal.pbio.0060159

Hassabis, D., Kumaran, D., Summerfield, C., and Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron* 95, 245–258. doi: 10.1016/j.neuron.2017.06.011

Hassabis, D., Kumaran, D., Vann, S. D., and Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1726–1731. doi: 10.1073/pnas.0610561104

Hastings, J., Frishkoff, G. A., Smith, B., Jensen, M., Poldrack, R. A., Lomax, J., et al. (2014). Interdisciplinary perspectives on the development, integration, and application of cognitive ontologies. *Front. Neuroinform.* 8:62. doi: 10.3389/fninf.2014.00062

He, Y., Wang, J., Wang, L., Chen, Z. J., Yan, C., Yang, H., et al. (2009). Uncovering intrinsic modular organization of spontaneous brain activity in humans. *PLoS One* 4, 23–25. doi: 10.1371/journal.pone.0005226

Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J. P., Meuli, R., et al. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2035–2040. doi: 10.1073/pnas.0811168106

Hutchins, L. N., Murphy, S. M., Singh, P., and Graber, J. H. (2008). Position-dependent motif characterization using non-negative matrix factorization. *Bioinformatics* 24, 2684–2690. doi: 10.1093/bioinformatics/btn526

Hwang, K., Bertolero, M. A., Liu, W. B., and D'Esposito, M. (2017). The human thalamus is an integrative hub for functional brain networks. *J. Neurosci.* 37, 5594–5607. doi: 10.1523/JNEUROSCI.0067-17.2017

Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nat. Rev. Neurosci.* 9, 304–313. doi: 10.1038/nrn2332

Jackson, R. L., Hoffman, P., Pobric, G., and Lambon Ralph, M. A. (2016). The semantic network at work and rest: differential connectivity of anterior temporal lobe subregions. *J. Neurosci.* 36, 1490–1501. doi: 10.1523/JNEUROSCI.2999-15.2016

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841. doi: 10.1016/S1053-8119(02)91132-8

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015

Jenkinson, M., and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–156. doi: 10.1016/S1361-8415(01)00036-6

Jonikaitis, D., and Moore, T. (2019). The interdependence of attention, working memory and gaze control: behavior and neural circuitry. *Curr. Opin. Psychol.* 29, 126–134. doi: 10.1016/j.copsyc.2019.01.012

Laird, A. R. (2009). ALE meta-analysis workflows via the BrainMap database: progress towards a probabilistic functional brain atlas. *Front. Neuroinform.* 3:23. doi: 10.3389/neuro.11.023.2009

Laird, A. R., Eickhoff, S. B., Fox, P. M., Uecker, A. M., Ray, K. L., Saenz, J. J., et al. (2011a). The BrainMap strategy for standardization, sharing, and meta-analysis of neuroimaging data. *BMC Res. Notes* 4:349. doi: 10.1186/1756-0500-4-349

Laird, A. R., Fox, P. M., Eickhoff, S. B., Turner, J. A., Ray, K. L., McKay, D. R., et al. (2011b). Behavioral interpretations of intrinsic connectivity networks. *J. Cogn. Neurosci.* 23, 4022–4037. doi: 10.1162/jocn_a_00077

Laird, A. R., Lancaster, J. L., and Fox, P. T. (2005). BrainMap: the social evolution of a human brain mapping database. *Neuroinformatics* 3, 65–78.

Lancaster, J. L., Tordesillas-Gutiérrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., et al. (2007). Bias between MNI and talairach coordinates analyzed using the ICBM-152 brain template. *Hum. Brain Mapp.* 28, 1194–1205. doi: 10.1002/hbm.20345

Lee, D. D., and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788–791. doi: 10.1038/44565

Lee, D. D., and Seung, H. S. (2001). Algorithms for non-negative matrix factorization. *Adv. Neural. Inf. Process. Syst.* 13, 556–562.

Lenartowicz, A., Kalar, D. J., Congdon, E., and Poldrack, R. A. (2010). Towards an ontology of cognitive control. *Top. Cogn. Sci.* 2, 678–692. doi: 10.1111/j.1756-8765.2010.01100.x

Li, J., Kong, R., Liegeois, R., Orban, C., Tan, Y., Sun, N., et al. (2019). Global signal regression strengthens association between resting-state functional connectivity and behavior. *NeuroImage* 196, 126–141. doi: 10.1016/j.neuroimage.2019.04.016

Li, W., Mai, X., and Liu, C. (2014). The default mode network and social understanding of others: what do brain connectivity studies tell us. *Front. Hum. Neurosci.* 8:74. doi: 10.3389/fnhum.2014.00074

Liu, J., Xia, M., Dai, Z., Wang, X., Liao, X., Bi, Y., et al. (2016). Intrinsic brain hub connectivity underlies individual differences in spatial working memory. *Cereb. Cortex* 27, 5496–5508. doi: 10.1093/cercor/bhw317

Lu, H., Zuo, Y., Gu, H., Waltz, J. A., Zhan, W., Scholl, C. A., et al. (2007). Synchronized delta oscillations correlate with the resting-state functional MRI signal. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18265–18269. doi: 10.1073/pnas.0705791104

Maass, W., Natschläger, T., and Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural. Comput.* 14, 2531–2560. doi: 10.1162/089976602760407955

Mars, R. B., Neubert, F., Noonan, M. P., Sallet, J., Toni, I., and Rushworth, M. F. S. (2012). On the relationship between the "default mode network" and the "social brain." *Front. Hum. Neurosci.* 6:189. doi: 10.3389/fnhum.2012.00189

Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., and Tononi, G. (2005). Breakdown of cortical effective connectivity during sleep. *Science* 309, 2228–2232. doi: 10.1126/science.1117256

Meyer, M. L., Davachi, L., Ochsner, K. N., and Lieberman, M. D. (2019). Evidence that default network connectivity during rest consolidates social information. *Cereb. Cortex* 29, 1910–1920. doi: 10.1093/cercor/bhy071

Mullally, S. L., and Maguire, E. A. (2014). Memory, imagination, and predicting the future: a common brain mechanism? *Neuroscientist* 20, 220–234. doi: 10.1177/1073858413495091

Murphy, K., Birn, R. M., Handwerker, D. A., Jones, T. B., and Bandettini, P. A. (2009). The impact of global signal regression on resting state correlations: are anti-correlated networks introduced?. *Neuroimage* 44, 893–905. doi: 10.1016/j.neuroimage.2008.09.036

Newman, M. (2010). *Networks: An Introduction*. New York, NY: Oxford University Press.

Nickel, M., and Kiela, D. (2017). Poincaré embeddings for learning hierarchical representations. *Adv. Neural. Inf. Process. Syst.* 30, 6338–6347.

Nigam, S., Shimono, M., Ito, S., Yeh, F. -C., Timme, N., Myroshnychenko, M., et al. (2016). Rich-club organization in effective connectivity among cortical neurons. *J. Neurosci.* 36, 670–684. doi: 10.1523/JNEUROSCI.2177-15.2016

Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435, 814–818. doi: 10.1038/nature03607

Passingham, R. E., and Wise, S. P. (2012). *The Neurobiology of the Prefrontal Cortex: Anatomy, Evolution, and the Origin of Insight*. Oxford: Oxford University Press.

Pauli, W. M., O'Reilly, R. C., Yarkoni, T., and Wager, T. D. (2016). Regional specialization within the human striatum for diverse psychological functions. *Proc. Natl. Acad. Sci. U.S.A.* 113, 1907–1912. doi: 10.1073/pnas.1507610113

Poldrack, R. A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., et al. (2011). The Cognitive Atlas: toward a knowledge foundation for cognitive neuroscience. *Front. Neuroinform.* 5:17. doi: 10.3389/fninf.2011.00017

Poldrack, R. A., Mumford, J. A., Schonberg, T., Kalar, D., Barman, B., and Yarkoni, T. (2012). Discovering relations between mind, brain, and mental disorders using topic mapping. *PLoS Comput. Biol.* 8:e1002707. doi: 10.1371/journal.pcbi.1002707

Poldrack, R. A., and Yarkoni, T. (2016). From brain maps to cognitive ontologies: informatics and the search for mental structure. *Annu. Rev. Psychol.* 67, 587–612. doi: 10.1146/annurev-psych-122414-033729

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 59, 2142–2154. doi: 10.1016/j.neuroimage.2011.10.018

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., et al. (2011). Functional network organization of the human brain. *Neuron* 72, 665–678. doi: 10.1016/j.neuron.2011.09.006

Power, J. D., Schlaggar, B. L., Lessov-Schlaggar, C. N., and Petersen, S. E. (2013). Evidence for hubs in human functional brain networks. *Neuron* 79, 798–813. doi: 10.1016/j.neuron.2013.07.035

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* 84, 320–341. doi: 10.1016/j.neuroimage.2013.08.048

Price, C. J., and Friston, K. J. (2005). Functional ontologies for cognition: the systematic definition of structure and function. *Cogn. Neuropsychol.* 22, 262–275. doi: 10.1080/02643290442000095

Proix, T., Spiegler, A., Schirner, M., Rothmeier, S., Ritter, P., and Jirsa, V. K. (2016). How do parcellation size and short-range connectivity affect dynamics in large-scale brain network models?. *NeuroImage* 142, 135–149. doi: 10.1016/j.neuroimage.2016.06.016

Ray, K. L., McKay, D. R., Fox, P. M., Riedel, M. C., Uecker, A. M., Beckmann, C. F., et al. (2013). ICA model order selection of task co-activation networks. *Front. Neurosci.* 7:237. doi: 10.3389/fnins.2013.00237

Reniers, R. L. E. P., Corcoran, R., Völlm, B. A., Mashru, A., Howard, R., and Liddle, P. F. (2012). Moral decision-making, ToM, empathy and the default mode network. *Biol. Psychol.* 90, 202–210. doi: 10.1016/j.biopsycho.2012.03.009

Rey-Mermet, A., Gade, M., Souza, A. S., von Bastian, C. C., and Oberauer, K. (2019). Is executive control related to working memory capacity and fluid intelligence? *J. Exp. Psychol. Gen.* 148, 1335–1372. doi: 10.1037/xge0000593

Roebroeck, A., Formisano, E., and Goebel, R. (2005). Mapping directed influence over the brain using Granger causality and fMRI. *Neuroimage* 25, 230–242. doi: 10.1016/j.neuroimage.2004.11.017

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65. doi: 10.1016/0377-0427(87)90125-7

Rubin, T. N., Koyejo, O., Gorgolewski, K. J., Jones, M. N., Poldrack, R. A., and Yarkoni, T. (2017). Decoding brain activity using a large-scale probabilistic functional-anatomical atlas of human cognition. *PLoS Comput. Biol.* 13:e1005649. doi: 10.1371/journal.pcbi.1005649

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., et al. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage* 64, 240–256. doi: 10.1016/j.neuroimage.2012.08.052

Schafer, M., and Schiller, D. (2018). Navigating social space. *Neuron* 100, 476–489. doi: 10.1016/j.neuron.2018.10.006

Schreiber, T. (2000). Measuring information transfer. *Phys. Rev. Lett.* 85, 461–464. doi: 10.1103/PhysRevLett.85.461

Seth, A. K. (2010). A MATLAB toolbox for Granger causal connectivity analysis. *J. Neurosci. Methods* 186, 262–273. doi: 10.1016/j.jneumeth.2009.11.020

Shen, X., Tokoglu, F., Papademetris, X., and Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *Neuroimage* 82, 403–415. doi: 10.1016/j.neuroimage.2013.05.081

Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155. doi: 10.1002/hbm.10062

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., et al. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13040–13045. doi: 10.1073/pnas.0905267106

Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., et al. (2013). Resting-state fMRI in the human connectome project. *Neuroimage* 80, 144–168. doi: 10.1016/j.neuroimage.2013.05.039

Sporns, O., Honey, C. J., and Kötter, R. (2007). Identification and classification of hubs in brain networks. *PLoS One* 2:e1049. doi: 10.1371/journal.pone.0001049

Sporns, O., and Kötter, R. (2004). Motifs in brain networks. *PLoS Biol.* 2:e369. doi: 10.1371/journal.pbio.0020369

Spreng, R. N., and Grady, C. L. (2010). Patterns of brain activity supporting autobiographical memory, prospection, and theory of mind, and their relationship to the default mode network. *J. Cogn. Neurosci.* 22, 1112–1123. doi: 10.1162/jocn.2009.21282

Spreng, R. N., Mar, R. A., and Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *J. Cogn. Neurosci.* 21, 489–510. doi: 10.1162/jocn.2008.21029

Stoodley, C. J. (2012). The cerebellum and cognition: evidence from functional imaging studies. *Cerebellum* 11, 352–365. doi: 10.1007/s12311-011-0260-7

Stoodley, C. J., and Schmahmann, J. D. (2009). Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *Neuroimage* 44, 489–501. doi: 10.1016/j.neuroimage.2008.08.039

Strick, P. L., Dum, R. P., and Fiez, J. A. (2009). Cerebellum and nonmotor function. *Annu. Rev. Neurosci.* 32, 413–434. doi: 10.1146/annurev.neuro.31.060407.125606

Tavares, R. M., Mendelsohn, A., Grossman, Y., Williams, C. H., Shapiro, M., Trope, Y., et al. (2015). A map for social navigation in the human brain. *Neuron* 87, 231–243. doi: 10.1016/j.neuron.2015.06.011

Turner, J. A., and Laird, A. R. (2012). The cognitive paradigm ontology: design and application. *Neuroinformatics* 10, 57–66. doi: 10.1007/s12021-011-9126-x

van den Heuvel, M. P., and Sporns, O. (2011). Rich-club organization of the human connectome. *J. Neurosci.* 31, 15775–15786. doi: 10.1523/JNEUROSCI.3539-11.2011

van den Heuvel, M. P., Stam, C. J., Boersma, M., and Hulshoff Pol, H. E. (2008). Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. *Neuroimage* 43, 528–539. doi: 10.1016/j.neuroimage.2008.08.010

van der Maaten, L. (2014). Accelerating t-SNE using tree-based algorithms. *J. Mach. Learn. Res.* 15, 3221–3245.

van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.

van Dijk, K. R. A., Sabuncu, M. R., and Buckner, R. L. (2012). The influence of head motion on intrinsic functional connectivity MRI. *Neuroimage* 59, 431–438. doi: 10.1016/j.neuroimage.2011.07.044

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., et al. (2013). The WU-Minn human connectome project: an overview. *Neuroimage* 80, 62–79. doi: 10.1016/j.neuroimage.2013.05.041

Varoquaux, G., Schwartz, Y., Poldrack, R. A., Gauthier, B., Bzdok, D., Poline, J. -B., et al. (2018). Atlases of cognition with large-scale human brain mapping. *PLoS Comput. Biol.* 14:e1006565. doi: 10.1371/journal.pcbi.1006565

Vicente, R., Wibral, M., Lindner, M., and Pipa, G. (2011). Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.* 30, 45–67. doi: 10.1007/s10827-010-0262-3

von Luxburg, U. (2007). A tutorial on spectral clustering. *Stat. Comput.* 17, 394–416. doi: 10.1007/s11222-007-9033-z

Wolpert, D. M., Miall, R. C., and Kawato, M. (1998). Internal models in the cerebellum. *Trends Cogn. Sci.* 2, 338–347. doi: 10.1016/S1364-6613(98)01221-2

Woolgar, A., Duncan, J., Manes, F., and Fedorenko, E. (2018). Fluid intelligence is supported by the multiple-demand system not the language system. *Nat. Hum. Behav.* 2, 200–204. doi: 10.1038/s41562-017-0282-3

Weissenbacher, A., Kasess, C., Gerstl, F., Lanzenberger, R., Moser, E., and Windischberger, C. (2009). Correlations and anticorrelations in resting-state functional connectivity MRI: a quantitative comparison of preprocessing strategies. *Neuroimage* 47, 1408–1416. doi: 10.1016/j.neuroimage.2009.05.005

Yamazaki, T., and Tanaka, S. (2007). The cerebellum as a liquid state machine. *Neural. Netw.* 20, 290–297. doi: 10.1016/j.neunet.2007.04.004

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* 8, 665–670. doi: 10.1038/nmeth.1635

Yu, S. X., and Shi, J. (2003). "Multiclass spectral clustering," in *Proceedings Ninth IEEE International Conference on Computer Vision*, Piscataway, NJ

Zhu, X., Zhu, Q., Jiang, C., Shen, H., Wang, F., Liao, W., et al. (2017). Disrupted resting-state default mode network in betel quid-dependent individuals. *Front. Psychol.* 8:84. doi: 10.3389/fpsyg.2017.00084

Žitnik, M., and Zupan, B. (2012). NIMFA: a Python library for nonnegative matrix factorization. *J. Mach. Learn. Res.* 13, 849–853.

Zuo, X. -N., Di Martino, A., Kelly, C., Shehzad, Z. E., Gee, D. G., Klein, D. F., et al. (2010). The oscillating brain: complex and reliable. *Neuroimage* 49, 1432–1445. doi: 10.1016/j.neuroimage.2009.09.037

Zuo, X., Ehmke, R., Mennes, M., Imperati, D., Castellanos, F. X., Sporns, O., et al. (2012). Network centrality in the human functional connectome. *Cereb. Cortex* 22, 1862–1875. doi: 10.1093/cercor/bhr269

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

frontiers in Human Neuroscience

BRIEF RESEARCH REPORT
published: 04 February 2020
doi: 10.3389/fnhum.2020.00010

Alpha Band Resting-State EEG Connectivity Is Associated With Non-verbal Intelligence

*Ilya Zakharov[1], Anna Tabueva[1], Timofey Adamovich[1], Yulia Kovas[2,3] and Sergey Malykh[1]**

[1] Developmental Behavioral Genetics Laboratory, Psychological Institute of the Russian Academy of Education, Moscow, Russia, [2] Department of Psychology, Goldsmiths University of London, London, United Kingdom, [3] International Centre for Research in Human Development, Tomsk State University, Tomsk, Russia

The aim of the present study was to investigate whether EEG resting state connectivity correlates with intelligence. One-hundred and sixty five participants took part in the study. Six minutes of eyes closed EEG resting state was recorded for each participant. Graph theoretical connectivity metrics were calculated separately for two well-established synchronization measures [weighted Phase Lag Index (wPLI) and Imaginary Coherence (iMCOH)] and for sensor- and source EEG space. Non-verbal intelligence was measured with Raven's Progressive Matrices. In line with the Neural Efficiency Hypothesis, path lengths characteristics of the brain networks (Average and Characteristic Path lengths, Diameter and Closeness Centrality) within alpha band range were significantly correlated with non-verbal intelligence for sensor space but no for source space. According to our results, variance in non-verbal intelligence measure can be mainly explained by the graph metrics built from the networks that include both weak and strong connections between the nodes.

Keywords: EEG, resting state, connectivity, intelligence, neural efficiency, graph theory

OPEN ACCESS

Edited by:
Chella Kamarajan, SUNY Downstate Medical Center, United States

Reviewed by:
Stavros I. Dimitriadis, Cardiff University, United Kingdom
Veena A. Nair, University of Wisconsin-Madison, United States

****Correspondence:***
Sergey Malykh
malykhsb@mail.ru

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *23 August 2019*
Accepted: *13 January 2020*
Published: *04 February 2020*

Citation:
Zakharov I, Tabueva A, Adamovich T, Kovas Y and Malykh S (2020) Alpha Band Resting-State EEG Connectivity Is Associated With Non-verbal Intelligence. Front. Hum. Neurosci. 14:10. doi: 10.3389/fnhum.2020.00010

INTRODUCTION

Information processing in the brain is reflected in brain oscillations (Ward, 2003; Buzsáki and Draguhn, 2004; Clayton et al., 2015; Sadaghiani and Kleinschmidt, 2016). However, it is still not clear how neurobiological factors actually contribute to more effective cognitive performance. One approach to understanding the relationship between brain functioning and cognition is the neural efficiency hypothesis of intelligence (Haier et al., 1988, 1992). According to this hypothesis, brains of more intelligent individuals work more efficiently when engaged in cognitive task performance as compared to those of less intelligent ones. In a seminal studies (Haier et al., 1988, 1992), using the Positron Emission Tomography (PET) method, participants with higher scores on Raven's progressive matrices were found to consume less glucose comparing to participants with lower scores. Later these results were extended to more types of brain activity measures (EEG, fMRI and so on) and different types of tasks (see Neubauer and Fink, 2009 for review).

The neural efficiency hypothesis predicts that the level of cognitive abilities would be correlated to brain activity during cognitive load. However, it is still unclear whether the brain activity at rest can be a good predictor of individual differences in intelligence. It has been proposed that the most informative way to investigate resting state activity is the network neuroscience approach

(Deary et al., 2010). This is because intelligence is not localized in a single area in the brain but rather operates through a distributed network (Bullmore and Sporns, 2009). According to this approach, the nervous system is a network of anatomically and functionally interconnected areas that form distinct functional systems operating in a coherent manner. The structure and dynamics of these complex systems and connectivity patterns within them can be studied with network modeling tools that originate from mathematical graph theory. "Efficiency" in case of network approach is defined in terms of the cost of transmitting information within the network. In particular, it appears that brain networks are organized in a way that achieves the maximum possible cost-efficiency: a topological structure that maximizes complexity while minimizing transmission costs (Bassett et al., 2010; Denève and Machens, 2016).

Connectivity patterns of the brain resting state activity have been shown to be highly stable for an individual (Finn et al., 2015) and could be used for prediction of various personality traits (e.g., temperament or creativity; Markett et al., 2013; Beaty et al., 2018) or associated with psychopathological states (see van den Heuvel and Sporns, 2019 for review). The growing consensus in this area of research includes several features of the topology of the brain networks important for intelligence: (1) neuronal ensembles incerebral cortex are organized into complex networks due to frequency specific oscillatory coupling (da Silva, 1991; Barahona and Pecora, 2002; Buzsaki, 2006; Senzai et al., 2019); (2) the brain network graphs of functional oscillatory activity patterns share cost-efficient "small-world" properties [meaning that there is only small number of steps from one node to any other (Sporns, 2007; Stam and van Straaten, 2012)]; (3) characteristics of frequency-specific networks architecture are unique for a person and can be used as the identifying "fingerprints" of the network with almost 100% accuracy (Finn et al., 2015; Yeh et al., 2016); (4) communication through neuronal coherence within neuronal networks represent the neural substrate for individual differences in cognitive processes (Fries, 2015).

However, data on the relationship between the brain resting state activity and individual level of intelligence is inconsistent. In some studies the brain resting state functional connectivity characteristics correlated to intelligence (Langer et al., 2012). However, a recent large-scale study of 1200 individuals from the Human Connectome Project failed to find any significant associations between measures of the brain resting state dynamics and several widely used intelligence measures (Kruschwitz et al., 2018). This lack of significant associations could be due to the method of assessment of functional connectivity. The study used fMRI BOLD signal oscillations which have poor temporal resolution (2–3 s, Logothetis, 2008). A number of studies have shown that brain oscillations of much higher frequency can play a significant role in cognition (Palva and Palva, 2011; Fries, 2015; Sockeel et al., 2016).

The aim of the present study was to replicate the association between graph metrics of EEG resting state brain connectivity and non-verbal intelligence, found by Langer et al. (2012); and to assess consistency of several widely used methods of calculating EEG connectivity.

MATERIALS AND METHODS

Graph connectivity metrics were assessed during resting state. In EEG, functional connectivity can be estimated for the oscillations directly recorded from electrodes (sensor space connectivity) or for the reconstructed sources of brain activity. In the present study we included graph connectivity metrics both for sensor and source EEG space. As signal estimates are spatially correlated, a leakage of electromagnetic activity into local source neighborhood often occurs. When the synchronization method ignores this effect, "false positive" findings typically arise. Various methods were proposed to overcome the spatial leakage problem (see Bastos and Schoffelen, 2016 for a review). In the present study we used two most popular measures designed to correct for spatial leakage to replicate our results: weighted Phase Lag Index (wPLI, Vinck et al., 2010) and Imaginary Coherence (iMCOH, Nolte et al., 2008).

To calculate graph connectivity metrics one has to choose the threshold synchronization value below which all the signal pairs are considered to be unrelated. In our study we systematically test several thresholds to understand how it affects the connectivity metrics and its relationship with non-verbal intelligence measure. The rationale for the network metrics choice and details of calculation are described in **Supplementary Table S1**. The plan of analysis is presented in **Figure 1**.

The non-verbal intelligence was measured with Raven's Standard Progressive Matrices (Raven and Court, 1998).

Participants

The participants were recruited via announcement in social networks (N = 165). They participated voluntarily without any monetary incentive. The exclusion criteria were any recorded history of psychiatric or neurological disorders and head trauma. Participants' age ranged from 17 to 34 (M = 21.7, SD = 3.36, 30% identified as female). The majority of the participants were students or had a bachelor degree.

Procedure

During resting state EEG acquisition all participants were instructed to sit still, think of nothing in particular and not to fall asleep for 10 min. Every 2 min the participants were asked to open or close their eyes with verbal instructions: "Now open your eyes," "Now close your eyes." Data with eyes closed were used for analysis in the present study.

The non-verbal intelligence was measured online before EEG recording with the shortened Raven's matrices test (Raven and Court, 1998). The test consists of series of incomplete matrices. In each task participants should choose one of the eight suggested variants to complete the pattern. The original test, comprises six sets – A, B, C, D, E, and F. Within each set, the 12 items progressively become more difficult. We used four sets – C, D, E, and F. Sets C, D, and E contained six item each: 1st, 3rd, 5th, 7th, 9th, and 11th (items with even numbers were excluded); and set F contained 12 items. Thus, there were 30 items in total. Sets were presented in the following order: C – >D – >E – >F, where each

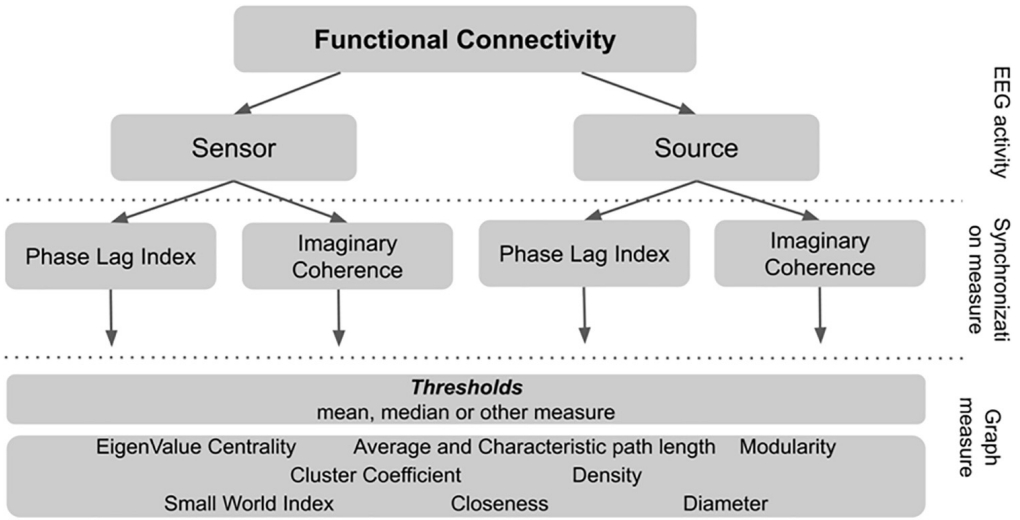


FIGURE 1 | The plan of analysis of the connectivity metrics calculation.

set in turn became more difficult. A total sum of correct items was used as the measure of general cognitive ability.

EEG Data Acquisition and Pre-processing

The EEG data was recorded from 64 active electrodes placed according to the international 10–10 system with Brain Products ActiChamp amplifier (BrainProducts, Munich, Germany). All experiments were conducted in a sound-attenuated and electrically shielded room with dim light. Impedance was kept under 25 kOhm with high conductive chloride gel. Approximate time for settling EEG was 15 min. The Brain Products PyCorder acquisition system was used for continuous recording without any filtering and continuously sampling at 500 Hz. The reference electrode was located at Cz. The data was re-referenced to the common reference after the recording and downsampled to 256 Hz. The data were filtered from 0.1 to 30 Hz and then re-referenced to an averaged reference and manually cleaned from artifacts, with noisy channels excluded. No more than 15% of the data was removed during artifact correction procedures. To remove blink and vertical eye-movement artifacts, independent component analysis (ICA) was performed on the following electrodes: VEOG — Fp1, HEOG — FT9 and FT10. After ICA, we topographically interpolated the excluded channels and conducted semiautomatic artifact rejection. The data were bandpassed into theta (4–8 Hz), alpha (8–13 Hz), beta1 (13–20 Hz), and beta2 (20–30 Hz) frequency ranges.

EEG Data Analysis

Synchronization Measures

To assess synchronization between pair of signals two metrics were used. Both metrics were calculated with MNE Python software (Gramfort et al., 2014).

Weighted Phase Lag Index (Vinck et al., 2010; Hardmeier et al., 2014) is an extension of the PLI, which quantifies the asymmetry of the relative phase distribution. PLI ignores amplitude and is robustto spurious increase in the coherence between signals due to common sources of brain activity.

$PLI = |\langle sig[\Delta\Phi(tk)]\rangle|$, where $\Delta\Phi(tk)$ – phase shift between two signals.

By weighing each phase difference according to the magnitude of the lag, phase differences around zero only marginally contribute to the calculation of the wPLI.

Imaginary Coherence (Nolte et al., 2008) – is another attempt to solve common source problem. The method is based on the assumption, that common source activity is reflected in different channels simultaneously, without time-lag. iMOCH is designed so that it is sensitive to time-lagged processes only.

The iMCOH could be calculated as:

$$icoh_{xy}(\omega) = \frac{Im(S_{xy}(\omega))}{\sqrt{S_{xx}(\omega)\,S_{yy}(\omega)}},$$

where $Im(S_{xy}(\omega))$ — part of the signal with time shift

Source Reconstruction

Source reconstruction was performed using standard source localization pipeline from MNE-package. First, source space with 503 sources for each hemisphere was created. Second, we used BEM (boundary-element model) to create three-layer model of the hemispheres. The three layers were inner skull, outer skull and outer skin. Conductivity of layers was standard for MNE package (0.3, 0.006, 0.3 for three layers accordingly). MNE exploit anatomical information from Free Surfer (Fuchs et al., 2001). Third, we constructed forward operator based on the source space and BEM model. Fourth, we created individual inverse operator for every participant with individual noise covariance matrix. Source reconstruction for each individual was performed

with appropriate inverse operator using dSPM method (Dale et al., 2000).

Connectivity Graph Measures

The connectivity metrics were chosen based on the reviews by Sporns and colleagues (Mišić and Sporns, 2016; Avena-Koenigsberger et al., 2018). In the present study we calculated the following graph connectivity metrics:

"Small world index" (SWI) – indexes the number of steps from one node to any other node within the network.
Average and Characteristic Path Length – the minimal number of edges that form a direct connection between two nodes (Average Path Length is based on the mean as the statistic, Characteristic Path Length – on the median).
Cluster Coefficient – a measure of the number of edges between a node's nearest neighbors or the fraction of triangles around a node, and is a measure of functional segregation. High *C* represents clustered connectivity at the node.
Modularity – a measure of functional segregation, which quantifies how well the network can be subdivided into non-overlapping groups of nodes or modules.
Diameter – the greatest distance between any pair of nodes within the network.
Eigenvector Centrality – a measure of the influence of a node in a network. A high *eigenvector* score means that a node is connected to many nodes that themselves have high scores.
Closeness Centrality – a measure of *centrality* in a network, calculated as the reciprocal of the sum of the length of the shortest paths between the node and all other nodes in the *graph*.
Graph measures were calculated with igraph package[1] for R (R Core Team, 2018).
The details of the calculation are presented in **Supplementary Table S1**.

STATISTICAL APPROACH

There is substantial variability in the possible routines of connectivity metrics calculation. There are several steps in EEG connectivity analysis where variability occurs. First, there is the alternative whether to use "raw" sensor-space EEG-signal or reconstruct and localize the source of EEG activity inside the brain. Second, there are different measures of synchronization between pairs of signals. Third, there is the convention to use only the pairs of signals with the strongest connections between them. However, the rationale to choose the "strong enough" threshold for synchronization measure is not explicitly and theoretically defined.

In our study we used two well-established but distinct measures of synchronization of the oscillatory brain activity (wPLI, Vinck et al., 2010; and iMCOH, Nolte et al., 2008). Synchronization was estimated for all pairs of EEG signals both for sensor and source EEG space separately for common EEG frequency bands (alpha, beta, and theta). In order to increase the number of investigated calculation alternatives and to decrease the number of multiple comparisons we developed the following approach used in the field of machine learning (Gareth, 2013). The whole sample was randomly divided into two subgroups: Test and Validation samples. Bootstrapped correlation coefficients were then separately calculated for the two samples– for non-verbal intelligence scores and all types of connectivity metrics. The median strength of connections within the person was used as the threshold (e.g., the 50% of the pairs with the highest synchronization estimates were used to calculate graph metrics). From this procedure we took only those metrics that were significantly correlated with non-verbal intelligence scores in both subsamples and both synchronization measures. To rule out possible impact of additional factors for these metrics, we also performed linear regression analysis with additional factors of sex and age of the participants. Another known variable, that can affect the EEG data is EEG spectral power. It was also added in the regression model.

The last step of the analysis was related to effect of thresholding on the connectivity metrics calculation. For the variables that remained significant after previous steps we calculated new metrics based on different synchronization thresholds (from 10 to 90% of the data preserved with the 10%-step). All the analyses were performed in R (R Core Team, 2018).

[1] http://igraph.org

RESULTS

EEG Sensor and Source Space Correlations

The consistent (repeated for different samples and synchronization measures) results were found only for alpha band EEG sensor space connectivity metrics. The bootstrapped correlations for the metrics in alpha band are presented in **Table 1**. The scatterplots for the relationship between wPLI-based metrics and intelligence are presented in **Figure 2** (The descriptive statistics and correlations for other frequency bands, as well as scatterplots for iMCOH-based metrics can be seen in **Supplementary Tables S2–S5** and **Supplementary Figures S1, S2**). There were a number of significant correlations between non-verbal intelligence and connectivity metrics for other frequency bands and EEG source space (see **Supplementary Tables S2–S4**), however, none of them were consistent for the different samples and synchronization measures.

At this step of analysis we were interested in the metrics that showed the same pattern of results both for wPLI and iMCOH measures and for both Test and Validation samples. Only four metrics calculated for alpha band EEG sensor space met these criteria (Average and Characteristic Path length, Diameter and Closeness). However, the collinearity analysis showed that multicollinearity was present for these metrics (VIF > 5, Ringle et al., 2015). Accordingly, for the next step of the analysis we used only one of these metrics. We have chosen Characteristic Path Length due to its most straightforward theoretical interpretation.

TABLE 1 | Bootstrapped correlations between alpha band connectivity metrics and non-verbal intelligence for test sample and validation sample.

Variable	EEG sensor space				EEG source space			
	WPLI		iMCOH		WPLI		iMCOH	
	Test	Validation	Test	Validation	Test	Validation	Test	Validation
Char PL	0.41**	0.31**	0.30**	0.32**	0.10	0.19	0.01	0.16
Average PL	0.43**	0.36**	0.30*	0.17	0.09	0.13	0.01	0.14
Clust coef.	−0.06	0.01	0.07	−0.28*	−0.00	−0.03	0.08	0.07
SWI	0.38**	0.27*	0.24	0.22	0.13	0.13	0.09	0.12
Modularity	−0.07	−0.20	−0.04	−0.09	−0.16	0.02	0.14	−0.18
Eigen. centrality	−0.06	0.16	−0.04	0.02	−0.02	−0.07	−0.05	0.01
Diameter	0.36**	0.29**	0.29*	0.29*	0.06	0.18	−0.00	0.16
Closeness	−0.35**	−0.28*	−0.38**	−0.31**	−0.17	−0.22	−0.09	−0.17

** indicates $p < 0.05$. ** indicates $p < 0.01$. Char PL, Characteristic Path Length; Average PL, Average Path Length; Clust Coef., Cluster coefficient; Eigen. Centrality, Eigenvector centrality.*

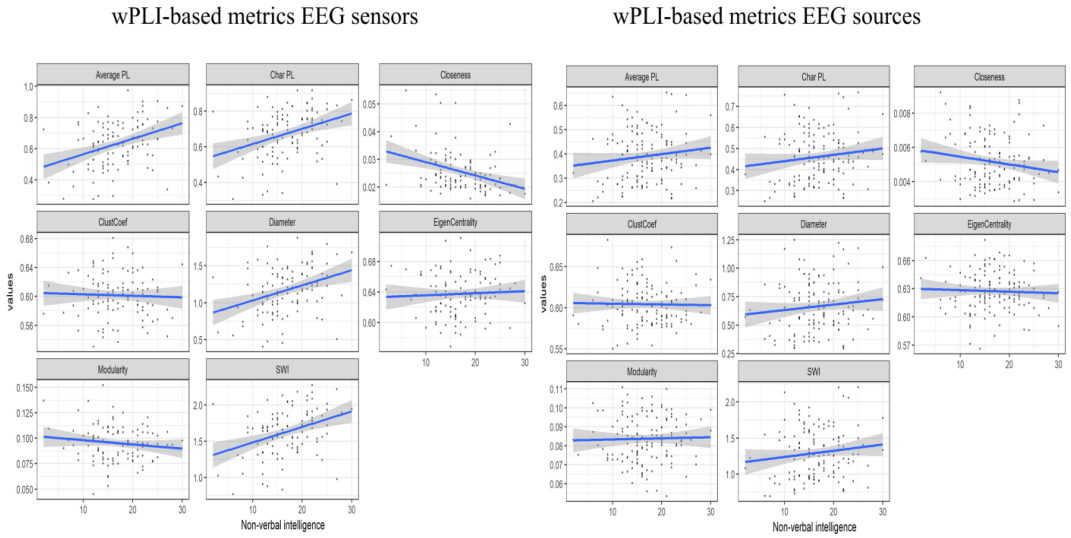


FIGURE 2 | The scatterplots for the relationship between the wPLI-based connectivity metrics and non-verbal intelligence.

The next step of the analysis was the linear regression analysis with sex, age and EEG spectral power as an additional variable. Both wPLI and iMCOH-based Characteristic Path Length remained statistically significant predictors of the level of non-verbal intelligence. The results are presented in **Tables 2**, **3**.

Correlations for Different EEG Sensor-Space Connectivity Matrix Construction Thresholds

One of the steps in the calculation of the connectivity metrics is thresholding procedure. Its main purpose is the increase in signal-to-noise ratio by deleting "weak" connections that do not contain any relevant physiological signal. The threshold values vary across studies considerably, which can lead to inconsistent results (Garrison et al., 2015). According to Sporns (2014) the average threshold is often chosen to delete the connections with the strength that lies at least below 75% of all connections. The discrepancies in the resulting networks with different percentiles (50th and 90th percentiles are taken as examples) can be seen in **Figure 3**.

The current study addressed the problem of choosing the threshold explicitly. We have calculated the sensor space Characteristic Path length with ten different thresholds (from 10 to 90% with the 10%-step) for the test sample. The results are presented in **Figure 3** (the detailed results are presented in **Supplementary Table S6**). No significant correlations with non-verbal intelligence were observed for the connectivity metrics calculated with 60% threshold or higher (i.e., with only strong connections between the nodes used to build the metric). The Characteristic Path Length metrics calculated with thresholds from 10 to 60% were significantly correlated with non-verbal intelligence, with r ranging from 0.24 to 0.36 ($p < 0.05$; adjusted for multiple comparisons with FDR correction). The pattern of these significant correlations is presented in **Figure 4**.

TABLE 2 | Regression results using non-verbal intelligence as the criterion and wPLI-based characteristic path length, sex, age, and EEG spectral power as predictors.

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit
(Intercept)	8.77*	[0.69, 16.85]						
Characteristic PL	14.86**	[8.11, 21.61]	0.39	[0.21, 0.57]	0.13	[0.02, 0.24]	0.38**	
sex	−1.51	[−3.38, 0.36]	−0.14	[−0.31, 0.03]	0.02	[−0.02, 0.06]	−0.13	
Full years	0.02	[−0.20, 0.23]	0.01	[−0.16, 0.18]	0.00	[−0.00, 0.00]	−0.04	
alpha_power	−0.00	[−0.07, 0.06]	−0.01	[−0.18, 0.17]	0.00	[−0.00, 0.00]	0.09	
								R^2 = 0.168**
								95% CI[0.04, 0.26]

*A significant b-weight indicates the beta-weight and semi-partial correlation are also significant. b represents unstandardized regression weights. beta indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. r represents the zero-order correlation. LL and UL indicate the lower and upper limits of a confidence interval, respectively. * indicates $p < 0.05$. ** indicates $p < 0.01$.*

TABLE 3 | Regression results using non-verbal intelligence as the criterion and iMCOH based characteristic path length, sex, age, and EEG spectral power as predictors.

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit
(Intercept)	14.80**	[8.15, 21.44]						
Characteristic PL	25.23**	[10.65, 39.81]	0.30	[0.13, 0.48]	0.08	[−0.01, 0.17]	0.32**	
sex	−1.16	[−3.08, 0.76]	−0.11	[−0.28, 0.07]	0.01	[−0.02, 0.04]	−0.12	
Full years	−0.01	[−0.21, 0.20]	−0.00	[−0.18, 0.17]	0.00	[−0.00, 0.00]	−0.02	
alpha_power	0.02	[−0.05, 0.08]	0.05	[−0.12, 0.23]	0.00	[−0.01, 0.02]	0.10	
								R^2 = 0.117**
								95% CI[0.01, 0.20]

*A significant b-weight indicates the beta-weight and semi-partial correlation are also significant. b represents unstandardized regression weights. beta indicates that the standardized regression weights. sr^2 represents the semi-partial correlation squared. r represents the zero-order correlation. LL and UL indicate the lower and upper limits of a confidence interval, respectively. ** indicates $p < 0.01$.*

DISCUSSION

The neural efficiency hypothesis of intelligence is an important example of the neuroscience-based theories that promote understanding of psychological phenomena. One of the promising and well-suited methods for testing the neural efficiency hypothesis is the graph theoretical approach to the brain network analysis. However, in the recent large-scale fMRI study with 1096 participants, the connectivity-derived metrics of the brain dynamics were shown to be uncorrelated with various intelligence measures. fMRI is limited in capturing high frequency oscillations. A suitable measure of fast brain activity related to cognition is EEG due to its high temporal resolution. The current study tested the hypothesis that EEG-derived connectivity metrics is associated with non-verbal intelligence.

We found that average and characteristic path lengths, the diameter, and closeness centrality of the network in alpha band in EEG sensor space are correlated with the level of non-verbal intelligence in both test and validation samples. In the initial study by Langer et al. (2012) the negative correlation between intelligence and characteristic path length of the networks has been found. In contrast, in our study the correlation between non-verbal intelligence and graph metrics, related to path lengths characteristics, were positive (except closeness centrality, which is in inverse relationship with path lengths). The path length measures of a network we used are thought to be related to the integrative capacity of individual elements and the entire network (Avena-Koenigsberger et al., 2018). The average and characteristic (median) path lengths are related to the ease of the information transferring in the network. The diameter of the network measures the length of the shortest path between the most distanced nodes of a graph. The higher the diameter, the less linked a network tends to be. In general, the brain networks are supposed to be organized to ensure reliable and efficient communication, while minimizing the spent resources. Networks with longer paths are supposed to be energetically and metabolically more costly to be developed (Bullmore and Sporns, 2012). At the same time, in real architectures near-minimal pathway structure violates strict mathematical minimization criteria and require additional energetic cost (Rubinov et al., 2015; Betzel et al., 2016). This fact is compatible with the notion, that nervous systems were evolutionary selected not only for cost minimization, but also for topological integration (Sousa et al., 2017; Ardesch et al., 2019).

The importance of the integration of different brain areas can also explain another result of our study. We have found that the weak connections in the EEG sensor space yielded the most important information about the relationship between connectivity metrics and non-verbal intelligence. In connectivity studies mostly strong connections between the nodes have been accounted for, while weak connections were eliminated from calculation of the metrics. But, according to our data, variance in intelligence measurecan be mainly explained by the graph metrics built from the networks with not only strong, but weak to moderate strength of connections between the nodes as well. Weak EEG connections could

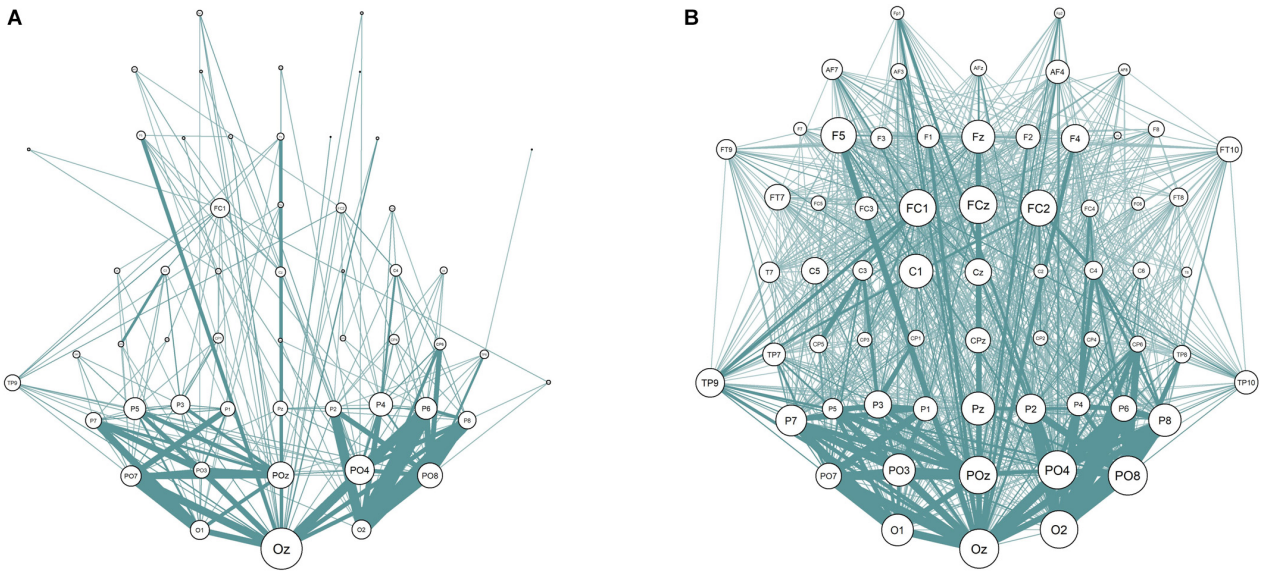


FIGURE 3 | EEG Sensor space synchronization networks for wPLI. Connections strength lower than 90% **(A)** and 50% **(B)** of all connections removed. Node size represents node degree; edge width represents strength of the connection between two nodes.

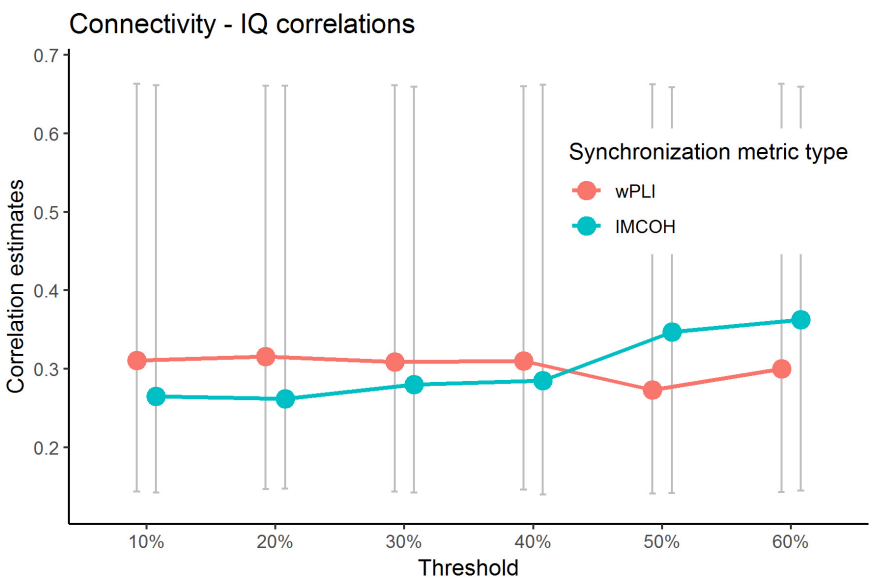


FIGURE 4 | Significant correlations between wPLI and iMCOH connectivity metrics and non-verbal intelligence for different thresholds. Only correlations with metrics built with thresholds of 60% or lower are presented ($p < 0.05$, adjusted for multiple comparisons with FDR correction). Non-significant correlations (with 70% and higher threshold) for Characteristic PL are not presented. The bars represent 95% confidence intervals for bootstrapped correlations.

represent additional abundant pathways that result in resilient communication within the network. Weak connections have been hypothesized to be important for stability of various types of networks (Granovetter, 1973; Pajevic and Plenz, 2012). Excessive number of pathways could be used to make the network less prone to bottlenecks and delays in the information transfer (Nassi and Callaway, 2009; Brandon et al., 2014. Several computational studies have demonstrated that geometrically embedded networks are, actually, characterized by physically short excessive connections that have an overall topology that increases the mean path length (Kaiser and Hilgetag, 2006). Research with macaque monkeys has shown that weak connections in the brainare important for neural cohesion (Goulas et al., 2015). Another possibility is that "a more intelligent brain" can engage larger amount of the distributed brain areas into the task solving process, which is in line with the parallel distribution processing theory (McClelland and Rumelhart, 1989; Bowers, 2017).

We have found that connectivity metrics of the brain networks oscillating within alpha frequency band range are sustainably correlated with non-verbal intelligence over wide variety of types of connectivity metrics, which is in line with the neural efficiency hypothesis. The EEG activity in the alpha range has been repeatedly associated with scores in various intelligence measures (Doppelmayr et al., 2002, 2005; Bazanova and Vernon, 2014). It has been hypothesized that the alpha band power reflects inhibition of non-essential processing (Klimesch et al., 2007). Thus, individual differences in the characteristics of resting-state alpha connectivity can indicate one's ability to inhibit activity irrelevant to the task at hand, which can lead to higher task performance at intelligence tests. The brain oscillations are supposed to be the mechanism that synchronize the activity between different areas. The frequency of oscillations has been shown to be related to the spatial scale of the synchronization. Rhythms with higher frequency synchronize communication on relatively small spatial scales within the brain, while slower rhythms synchronize activity between more distant brain areas (Buzsaki, 2006; Fries, 2015). Therefore, the large-scale communication between the brain areas during resting state can be one of the key mechanisms underlying individual differences in cognitive functioning.

Our results are in line with recent diffusion tensor imaging study, where the preserving of weak connections on calculation of graph connectivity metrics was advocated for Civier et al. (2019), and with earlier fMRI study where thresholding has been shown to lead to inconsistent results (Garrison et al., 2015). The results suggest that common practice in research to eliminate weak connections may lead to missing important information. One way to avoid arbitrary thresholding is the data-driven approach to construction of the graphs, which can be based, for example, on the minimal spanning tree algorithms, (Dimitriadis et al., 2017, 2018). However, spanning-tree approach can be highly dependent on an initial state of the data and lead to different results with minor changes in the signal. Thus, the direct comparison of the results obtained with thresholding approaches and various data-driven approaches are needed.

Our study has a number of limitations. First, there is no consensus in the literature about the best measure of EEG synchronization, and the most stable source localization approach. In the current study we have chosen two of the most favorites synchronization measures and one paradigm of EEG source localization. Further research is needed to test whether our results can be replicated with other types of measures: (1) synchronization- orthogonalized envelope correlation; canonical coherence; bicoherence; phase shift invariant imaging of coherent sources (Ossadtchi et al., 2018,Vidaurre et al., 2019); and (2) localization – LORETA, FOCUSS, MUSIC (Jatoi et al., 2014).

Second, we hypothesize that the discrepancy in the association between intelligence level and EEG and fMRI-derived connectivity metrics can be attributed to higher EEG temporal resolution. In the present study we were not able to compare these methods directly. The combined EEG-fMRI study is needed to further investigate this question.

Lastly, in the initial studies, the neural efficiency hypothesis was based on the analysis of the brain activity during cognitive performance rather than resting state. To clarify whether the graph theoretical approach is a good measure of efficiency within neural networks under cognitive load, a replication of classical studies with the graph connectivity metrics calculation is needed.

Overall, we have found that the connectivity characteristics of the brain networks (particularly, oscillating within alpha range), derived from EEG resting state with graph theoretical approach, are significantly correlated with non-verbal intelligence. This is in line with the neural efficiency hypothesis. According to our results, weak EEG connections contain an important information about the brain activity. Therefore, it is possible that the widely used thresholding procedure can lead to increase in false negative results.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

This study received approval from the Ethics Committee of the Psychological Institute of the Russian Academy of Education. For all participants written informed consent was obtained.

AUTHOR CONTRIBUTIONS

IZ, AT, and SM contributed to the conception and design of the study. IZ and AT collected the data. IZ, AT, and TA performed the statistical analysis. IZ wrote the first draft of the manuscript. AT and TA assisted in writting the manuscript. YK and SM edited and finally approved the submitted manuscript. All authors read and approved of the submitted version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00010/full#supplementary-material

REFERENCES

Ardesch, D. J., Scholtens, L. H., Li, L., Preuss, T. M., Rilling, J. K., and van den Heuvel, M. P. (2019). Evolutionary expansion of connectivity between multimodal association areas in the human brain compared with chimpanzees. *Proc. Natl. Acad. Sci. U.S.A.* 116, 7101–7106. doi: 10.1073/pnas.1818512116

Avena-Koenigsberger, A., Misic, B., and Sporns, O. (2018). Communication dynamics in complex brain networks. *Nat. Rev. Neurosc.* 19:17.

Barahona, M., and Pecora, L. M. (2002). Synchronization in small-world systems. *Phys. Rev. Lett.* 89:054101.

Bassett, D. S., Greenfield, D. L., Meyer-Lindenberg, A., Weinberger, D. R., Moore, S. W., and Bullmore, E. T. (2010). Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. *PLoS Comput. Biol.* 6:e1000748. doi: 10.1371/journal.pcbi.1000748

Bastos, A. M., and Schoffelen, J.-M. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Front. Syst. Neurosci.* 9:175. doi: 10.3389/fnsys.2015.00175

Bazanova, O. M., and Vernon, D. (2014). Interpreting EEG alpha activity. *Neurosci. Biobehav. Rev.* 44, 94–110. doi: 10.1016/j.neubiorev.2013.05.007

Beaty, R. E., Kenett, Y. N., Christensen, A. P., Rosenberg, M. D., Benedek, M., Chen, Q., et al. (2018). Robust prediction of individual creative ability from brain functional connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 115, 1087–1092. doi: 10.1073/pnas.1713532115

Betzel, R., Avena-Koenigsberger, A., Goñi, J., He, Y., de Reus, M. A., Griffa, A. F., et al. (2016). Generative models of the human connectome. *Neuroimage* 124, 1054–1064. doi: 10.1016/j.neuroimage.2015.09.041

Bowers, J. S. (2017). Parallel distributed processing theory in the age of deep networks. *Trends Cogn. Sci.* 21, 950–961. doi: 10.1016/j.tics.2017.09.013

Brandon, M. P., Koenig, J., and Leutgeb, S. (2014). Parallel and convergent processing in grid cell, head-direction cell, boundary cell, and place cell networks. *Wiley Interdiscipl. Rev. Cogn. Sci.* 5, 207–219. doi: 10.1002/wcs.1272

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10:186. doi: 10.1038/nrn2575

Bullmore, E., and Sporns, O. (2012). The economy of brain network organization. *Nat. Rev. Neurosci.* 13, 336–349. doi: 10.1038/nrn3214

Buzsaki, G. (2006). *Rhythms of the Brain*. Oxford: Oxford University Press.

Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi: 10.1126/science.1099745

Civier, O., Smith, R. E., Yeh, C. H., Connelly, A., and Calamante, F. (2019). Is removal of weak connections necessary for graph-theoretical analysis of dense weighted structural connectomes from diffusion MRI?. *Neuroimage* 194, 68–81. doi: 10.1016/j.neuroimage.2019.02.039

Clayton, M. S., Yeung, N., and Kadosh, R. C. (2015). The roles of cortical oscillations in sustained attention. *Trends Cogn. Sci.* 19, 188–195. doi: 10.1016/j.tics.2015.02.004

da Silva, F. L. (1991). Neural mechanisms underlying brain waves: from neural membranes to networks. *Electroencephalogr. Clin. Neurophysiol.* 79, 81–93. doi: 10.1016/0013-4694(91)90044-5

Dale, A., Liu, A., Fischl, B., and Buckner, R. (2000). Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26, 55–67.

Deary, I. J., Penke, L., and Johnson, W. (2010). The neuroscience of human intelligence differences. *Nat. Rev. Neurosci.* 11:201. doi: 10.1038/nrn2793

Denève, S., and Machens, C. K. (2016). Efficient codes and balanced networks. *Nat. Neurosci.* 19, 375–382. doi: 10.1038/nn.4243

Dimitriadis, S. I., Antonakakis, M., Simos, P., Fletcher, J. M., and Papanicolaou, A. C. (2018). Data-driven topological filtering based on orthogonal minimal spanning trees: application to multigroupmagnetoencephalography resting-state connectivity. *Brain Connect.* 7, 661–670. doi: 10.1089/brain.2017.0512

Dimitriadis, S. I., Salis, C., Tarnanas, I., and Linden, D. E. (2017). Topological filtering of dynamic functional brain networks unfolds informative chronnectomics: a novel data-driven thresholding scheme based on orthogonal minimal spanning trees (OMSTs). *Front. Neuroinform.* 11:28. doi: 10.3389/fninf.2017.00028

Doppelmayr, M., Klimesch, W., Sauseng, P., Hödlmoser, K., Stadler, W., and Hanslmayr, S. (2005). Intelligence related differences in EEG-bandpower. *Neurosci. Lett.* 381, 309–313. doi: 10.1016/j.neulet.2005.02.037

Doppelmayr, M., Klimesch, W., Stadler, W., Pöllhuber, D., and Heine, C. (2002). EEG alpha power and intelligence. *Intelligence* 30, 289–302. doi: 10.1016/s0160-2896(01)00101-5

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., et al. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* 18, 1664–1671. doi: 10.1038/nn.4135

Fries, P. (2015). Rhythms for cognition: communication through coherence. *Neuron* 88, 220–235. doi: 10.1016/j.neuron.2015.09.034

Fuchs, M., Wagner, M., and Kastner, J. (2001). Boundary element method volume conductor models for EEG source reconstruction. *Clin. Neurophysiol.* 112, 1400–1407. doi: 10.1016/s1388-2457(01)00589-2

Gareth, J. (2013). *An Introduction to Statistical Learning: with Applications in R*. Berlin: Springer.

Garrison, K. A., Scheinost, D., Finn, E. S., Shen, X., and Constable, R. T. (2015). The (in) stability of functional brain network measures across thresholds. *Neuroimage* 118, 651–661. doi: 10.1016/j.neuroimage.2015.05.046

Goulas, A., Schaefer, A., and Margulies, D. S. (2015). The strength of weak connections in the macaque cortico-cortical network. *Brain Struct. Funct.* 220, 2939–2951. doi: 10.1007/s00429-014-0836-3

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2014). MNE software for processing MEG and EEG data. *Neuroimage* 86, 446–460. doi: 10.1016/j.neuroimage.2013.10.027

Granovetter, M. S. (1973). The strength of weak ties. *Am. J. Sociol.* 78, 1360–1380.

Haier, R. J., Siegel, B., Tang, C., Abel, L., and Buchsbaum, M. S. (1992). Intelligence and changes in regional cerebral glucose metabolic rate following learning. *Intelligence* 16, 415–426. doi: 10.1016/0160-2896(92)90018-m

Haier, R. J., Siegel, B. V. Jr., Nuechterlein, K. H., Hazlett, E., Wu, J. C., Paek, J., et al. (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence* 12, 199–217. doi: 10.1016/j.neubiorev.2009.04.001

Hardmeier, M., Hatz, F., Bousleiman, H., Schindler, C., Stam, C. J., and Fuhr, P. (2014). Reproducibility of functional connectivity and graph measures based on the phase lag index (PLI) and weighted phase lag index (wPLI) derived from high resolution EEG. *PLoS One* 9:e108648. doi: 10.1371/journal.pone.0108648

Jatoi, M. A., Kamel, N., Malik, A. S., Faye, I., and Begum, T. (2014). A survey of methods used for source localization using EEG signals. *Biomed. Signal Proces. Control* 11, 42–52. doi: 10.1016/j.bspc.2014.01.009

Kaiser, M., and Hilgetag, C. C. (2006). Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS Comput. Biol.* 2:e95. doi: 10.1371/journal.pcbi.0020095

Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition–timing hypothesis. *Brain Res. Rev.* 53, 63–88. doi: 10.1016/j.brainresrev.2006.06.003

Kruschwitz, J. D., Waller, L., Daedelow, L. S., Walter, H., and Veer, I. M. (2018). General, crystallized and fluid intelligence are not associated with functional global network efficiency: a replication study with the human connectome project 1200 data set. *Neuroimage* 171, 323–331. doi: 10.1016/j.neuroimage.2018.01.018

Langer, N., Pedroni, A., Gianotti, L. R. R., Hänggi, J., Knoch, D., and Jäncke, L. (2012). Functional brain network efficiency predicts intelligence. *Hum. Brain Map.* 33, 1393–1406. doi: 10.1002/hbm.21297

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature* 453:869. doi: 10.1038/nature06976

Markett, S., Weber, B., Voigt, G., Montag, C., Felten, A., and Elger, C. (2013). Intrinsic connectivity networks and personality: the temperament dimension harm avoidance moderates functional connectivity in the resting brain. *Neuroscience* 240, 98–105. doi: 10.1016/j.neuroscience.2013.02.056

McClelland, J. L., and Rumelhart, D. E. (1989). *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises*. Massachusetts, US: MIT press.

Mišić, B., and Sporns, O. (2016). From regions to connections and networks: new bridges between brain and behavior. *Curr. Opin. Neurobiol.* 40, 1–7. doi: 10.1016/j.conb.2016.05.003

Nassi, J. J., and Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. *Nat. Rev. Neurosci.* 10:360. doi: 10.1038/nrn2619

Neubauer, A. C., and Fink, A. (2009). Intelligence and neural efficiency: measures of brain activation versus measures of functional connectivity in the brain. *Intelligence* 37, 223–229. doi: 10.1016/j.intell.2008.10.008

Nolte, G., Ziehe, A., Nikulin, V. V., Schlögl, A., Krämer, N., Brismar, T., et al. (2008). Robustly estimating the flow direction of information in complex physical systems. *Phys. Rev. Lett.* 100:234101. doi: 10.1103/PhysRevLett.100.234101

Ossadtchi, A., Altukhov, D., and Jerbi, K. (2018). Phase shift invariant imaging of coherent sources (PSIICOS) from MEG data. *Neuroimage* 183, 950–971. doi: 10.1016/j.neuroimage.2018.08.031

Pajevic, S., and Plenz, D. (2012). The organization of strong links in complex networks. *Nat. Phys.* 8:429. doi: 10.1038/nphys2257

Palva, S., and Palva, J. M. (2011). Functional roles of alpha-band phase synchronization in local and large-scale cortical networks. *Front. Psychol.* 2:204. doi: 10.3389/fpsyg.2011.00204

R Core Team, (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Raven, J. C., and Court, J. H. (1998). *Raven's Progressive Matrices and Vocabulary Scales*. Oxford: Oxford pyschologists Press.

Ringle, C. M., Wende, S., and Becker, J. M. (2015). *SmartPLS 3*. Bönningstedt: SmartPLS,

Rubinov, M., Ypma, R. J. F., Watson, C., and Bullmore, E. T. (2015). Wiring cost and topological participation of the mouse brain connectome. *Proc. Natl Acad. Sci. U.S.A.* 112, 10032–10037. doi: 10.1073/pnas.1420315112

Sadaghiani, S., and Kleinschmidt, A. (2016). Brain networks and α-oscillations: structural and functional foundations of cognitive control. *Trends Cogn. Sci.* 20, 805–817. doi: 10.1016/j.tics.2016.09.004

Senzai, Y., Fernandez-Ruiz, A., and Buzsáki, G. (2019). Layer-specific physiological features and interlaminar interactions in the primary visual cortex of the mouse. *Neuron* 101, 500–513. doi: 10.1016/j.neuron.2018.12.009

Sockeel, S., Schwartz, D., Pélégrini-Issac, M., and Benali, H. (2016). Large-scale functional networks identified from resting-state EEG using spatial ICA. *PLoS One* 11:e0146845. doi: 10.1371/journal.pone.0146845

Sousa, A. M., Meyer, K. A., Santpere, G., Gulden, F. O., and Sestan, N. (2017). Evolution of the human nervous system function, structure, and development. *Cell* 170, 226–247. doi: 10.1016/j.cell.2017.06.036

Sporns, O. (2007). Brain connectivity. *Scholarpedia* 2:4695. doi: 10.4249/scholarpedia.4695

Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience. *Nat. Neurosci.* 17, 652–660. doi: 10.1038/nn.3690

Stam, C. V., and van Straaten, E. C. W. (2012). The organization of physiological brain networks. *Clin. Neurophysiol.* 123, 1067–1087. doi: 10.1016/j.clinph.2012.01.011

van den Heuvel, M. P., and Sporns, O. (2019). A cross-disorder connectome landscape of brain dysconnectivity. *Nat. Rev. Neurosci.* 20, 435–446 doi: 10.1038/s41583-019-0177-6

Vidaurre, C., Nolte, G., de Vries, I. E. J., Gómez, M., Boonstra, T. W., Müller, K. R., et al. (2019). Canonical maximization of coherence: a novel tool for investigation of neuronal interactions between two datasets. *Neuroimage* 201:116009. doi: 10.1016/j.neuroimage.2019.116009

Vinck, M., van Wingerden, M., Womelsdorf, T., Fries, P., and Pennartz, C. M. A. (2010). The pairwise phase consistency: a bias-free measure of rhythmic neuronal synchronization. *Neuroimage* 51, 112–122. doi: 10.1016/j.neuroimage.2010.01.073

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends Cogn. Sci.* 7, 553–559. doi: 10.1016/j.tics.2003.10.012

Yeh, F. C., Vettel, J. M., Singh, A., Poczos, B., Grafton, S. T., Erickson, K. I, et al. (2016). Quantifying differences and similarities in whole-brain white matter architecture using local connectome fingerprints. *PLoS Comput. Biol.* 12:e1005203. doi: 10.1371/journal.pcbi.1005203

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

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 27 March 2020
doi: 10.3389/fnhum.2020.00116

Associations Between Individual Differences in Mathematical Competencies and Surface Anatomy of the Adult Brain

Alexander E. Heidekum, Stephan E. Vogel and Roland H. Grabner*

Educational Neuroscience, Institute of Psychology, University of Graz, Graz, Austria

Previously conducted structural magnetic resonance imaging (MRI) studies on the neuroanatomical correlates of mathematical abilities and competencies have several methodological limitations. Besides small sample sizes, the majority of these studies have employed voxel-based morphometry (VBM)—a method that, although it is easy to implement, has some major drawbacks. Taking this into account, the current study is the first to investigate in a large sample of typically developed adults the associations between mathematical abilities and variations in brain surface structure by using surface-based morphometry (SBM). SBM is a method that also allows the investigation of brain morphometry by avoiding the pitfalls of VBM. Eighty-nine young adults were tested with a large battery of psychometric tests to measure mathematical competencies in four different areas: (1) simple arithmetic; (2) complex arithmetic; (3) higher-order mathematics; and (4) numerical intelligence. Also, we asked participants for their mathematics grades for their final school exams. Inside the MRI scanner, we collected high-resolution T1-weighted anatomical images from each subject. SBM analyses were performed with the computational anatomy toolbox (CAT12) and indices for cortical thickness, for cortical surface complexity, for gyrification, and sulcal depth were calculated. Further analyses revealed associations between: (1) the cortical surface complexity of the right superior temporal gyrus and numerical intelligence; (2) the depth of the right central sulcus and adults' ability to solve complex arithmetic problems; and (3) the depth of the left parieto-occipital sulcus and adults' higher-order mathematics competence. Interestingly, no relationships with previously reported brain regions were observed, thus, suggesting the importance of similar research to confirm the role of the brain regions found in this study.

Keywords: right superior temporal gyrus, right central sulcus, left parieto-occipital sulcus, surface-based morphometry, mathematics

OPEN ACCESS

Edited by:
Elise Klein,
University of Tübingen, Germany

Reviewed by:
Tanya M. Evans,
University of Virginia, United States
Mario Pinto,
Santa Lucia Foundation (IRCCS), Italy

****Correspondence:***
Alexander E. Heidekum
alexander.heidekum@uni-graz.at

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *15 October 2019*
Accepted: *13 March 2020*
Published: *27 March 2020*

Citation:
Heidekum AE, Vogel SE and Grabner RH (2020) Associations Between Individual Differences in Mathematical Competencies and Surface Anatomy of the Adult Brain. Front. Hum. Neurosci. 14:116. doi: 10.3389/fnhum.2020.00116

INTRODUCTION

Mathematical competencies are one of the key cognitive abilities in our modern societies, and they are crucial for our profession as well as social development (Parsons and Bynner, 2005). To date, an extensive number of neuroimaging studies have investigated the cognitive architecture of mathematical cognition. However, these studies have mainly focused on neurofunctional aspects of mathematical competencies, whereas literature on their neuroanatomical correlates is scarce.

Especially, the surface characteristics of the gray matter (GM; i.e., cortical surface complexity, cortical thickness, gyrification, and sulcal depth) and their relationships with mathematical competencies have been widely ignored. The current study aimed to shed light on this aspect.

Numerous functional magnetic resonance imaging (fMRI) studies have demonstrated that numerical and mathematical cognition is associated with the activation of various frontal (e.g., ventral- and dorsolateral prefrontal cortex) and parietal brain regions (e.g., inferior and superior parietal cortex; for a review see Menon, 2014; for a meta-analysis see Arsalidou and Taylor, 2011). Also, it has been shown that the recruitment of these brain regions depends on the involved cognitive processes (e.g., quantity processing, arithmetic problem solving, etc.; Dehaene et al., 2003; Arsalidou and Taylor, 2011). For instance, while activation within the intraparietal sulcal (IPS) is thought to reflect quantity-based processes (Dehaene et al., 2003; Wilkey et al., 2017; Vogel et al., 2017a; or for a recent meta-analysis see Sokolowski et al., 2017), the controlled retrieval of a solution to a given arithmetic problem seems to rely on brain regions such as the inferior frontal gyrus (IFG), the supramarginal gyrus (SMG) and the angular gyrus (AG; Delazer et al., 2003, 2005; Ischebeck et al., 2006, 2009; Grabner et al., 2009; Klein et al., 2013, 2014; Menon, 2014; Peters and De Smedt, 2018; Heidekum et al., 2019). Furthermore, neuroimaging data indicate that activity in this network is also modulated by individual differences (e.g., Grabner et al., 2007; De Smedt and Gilmore, 2011; Berteletti et al., 2014). For instance, Grabner et al. (2007) compared two groups of young adults with different levels of mathematical competence who solved single-digit and multi-digit multiplication problems. The authors showed that individuals with higher mathematical competence more strongly activate the left AG while solving easy and more difficult multiplication problems. Grabner et al. (2007) concluded that mathematically more (compared to less) competent individuals rely more strongly on language-mediated processes (in particular fact retrieval) during arithmetic problem-solving.

In contrast to the large number of neurofunctional studies, only a few studies have used high-resolution structural MRI to examine how individual differences in brain morphology relate to variations in mathematical abilities and competencies (e.g., Isaacs et al., 2001; Aydin et al., 2007; Han et al., 2008, 2013; Rotzer et al., 2008; Rykhlevskaia et al., 2009; Lubin et al., 2013; Ranpura et al., 2013; Starke et al., 2013; Supekar et al., 2013; Cappelletti and Price, 2014; Evans et al., 2015; Price et al., 2016; Wilkey et al., 2018; Moreau et al., 2019). Many of those studies have focused on the comparison between dyscalculic individuals and controls (Han et al., 2008; Rotzer et al., 2008; Rykhlevskaia et al., 2009; Ranpura et al., 2013; Starke et al., 2013; Cappelletti and Price, 2014). For instance, it has been shown that dyscalculic adults have less GM volume in the right parietal cortex compared to controls (Cappelletti and Price, 2014) and that children with low mathematical skills relative to gender and age-matched controls show structural differences in the bilateral parietal lobes, right frontal lobe, and left occipital/parietal lobe (Han et al., 2008). These findings suggest an overlap of brain regions found in structural MRI studies and those reported in functional neuroimaging studies.

There are only a few studies (Supekar et al., 2013; Price et al., 2016; Wilkey et al., 2018) that investigated the neuroanatomical correlates of mathematics performance in typically developed individuals. However, most of these studies were conducted with children. For instance, Price et al. (2016) assessed the longitudinal and concurrent relations between GM volume and mathematical performance in a group of 50 children. GM volume maps derived from anatomical scans collected at the end of 1st grade and 2nd grade were related to performance on a standardized math test (i.e., Woodcock-Johnson III Tests of Achievement Calculation and Applied Problems). They found that left IPS GM volume at the end of 1st grade was positively related to math competence at the end of 2nd grade. Additionally, a positive association between GM volume in the same brain region at the end of 2nd grade and children's concurrent math competence was observed. Wilkey et al. (2018) measured children's performance in grade-level mathematics by a state-wide, school-based test of math achievement. They reported positive associations between differences in GM volume of the left and right hippocampal formations (including the hippocampus proper, entorhinal cortex, and subiculum) as well as the right IFG and children's performance in grade-level mathematics.

Previous structural MRI studies on the neuroanatomical correlates of mathematical abilities and competencies have several methodological limitations. As highlighted above, the large majority of studies have involved children with mathematical learning disabilities (e.g., Han et al., 2013; Starke et al., 2013). Thus, they do not allow reliable inferences to typically developed populations because we cannot assume that neuroanatomical mechanisms that distinguish dyscalculic individuals from controls are the same that underlie individual differences in (typically developing) mathematical competence. Furthermore, we do not know whether the neuroanatomical mechanisms important during development are still playing a role in adults. Another methodological problem is the number of participants that were under investigation. Many findings are based on small sample sizes (Rotzer et al., 2008; Han et al., 2013), which is associated with various problems, such as low statistical power (Button et al., 2013) or an inflated false discovery rate (Colquhoun, 2014). Finally, most of these studies employed voxel-based morphometry (VBM) to investigate relationships between GM density and mathematical competencies. VBM is a technique that remains extremely popular because it is highly automated and therefore quick and easy to use. However, VBM also has been associated with important limitations in recent years. For example, recent studies have shown that differences in methodological choices like the registration algorithm (Peelle et al., 2012), or changes of user-specified parameters such as the smoothing kernel size (Henley et al., 2010) can lead to inconsistent VBM results. Furthermore, the interpretation of VBM findings is often difficult because the results may be driven by differences in cortical thickness, surface area (SA), cortical volume, and folding or any combination of these measures (Voets et al., 2008; Hutton et al., 2009).

A currently very promising alternative to VBM is surface-based morphometry (SBM), which also allows the investigation of brain morphometry. However, SBM uses brain surface meshes for spatial registration, which increases the accuracy of brain registration compared to mere volume-based registration (Desai et al., 2005). Further, it allows the calculation of additional measures of the neocortical surface structure, namely: cortical thickness (Dahnke et al., 2013), cortical surface complexity (quantification of the spatial frequency of gyrification and fissuration of the brain surface; Yotter et al., 2011), gyrification (Luders et al., 2006) and sulcal depth (Yotter et al., 2011). Due to these and further advantages, SBM has been frequently used in recent years, and a growing body of evidence suggests that cortical surface measures are particularly informative for individual differences in intellectual abilities (e.g., Geschwind and Rakic, 2013). For instance, longitudinal changes in cortical thickness have been related to variations in intelligence in children (Schnack et al., 2015) and a pronounced thickening of specific areas of the neocortex has been related to intellectual abilities in adulthood (Brans et al., 2010).

Only one study has used SBM in mathematical cognition so far. Moreau et al. (2019) investigated relationships between volumetric and surface characteristics of GM and dyslexia, dyscalculia and comorbid manifestation of both. They collected MRI data from four different groups of adults (i.e., dyslexic, dyscalculic, comorbid and control) and performed VBM as well as SBM analyses. By using Bayesian methods, Moreau et al. (2019) did not find any evidence for group differences in GM volume or any surface characteristics (i.e., cortical thickness, gyrification, sulcal depth, or cortical complexity). Therefore, the authors concluded that GM differences associated with these developmental disorders are not as reliable as previously suggested. However, these results were based on relatively small group samples ($N = 12$) and do not allow inferences to individual differences in typically developed populations.

Against this background, the current study aimed to investigate how structural variations in GM characteristics relate to individual differences in mathematical competencies in adults. Limitations of previous work were taken into account when designing the current study. First of all, to increase the power of the current study and the generalizability of its results, we collected data from a comparably large adult sample consisting of typically developed individuals. Thus, we attempted to identify neuroanatomical correlates of mathematical competencies in a population that has been underrepresented in previous studies. This will provide important insights into how the human brain mediates individual differences in mathematical cognition. Second, since previous studies showed that the neural correlates of mathematical competence depend on the mathematical demand (Price et al., 2016; Wilkey et al., 2018), we examined a broad range of mathematical competencies (e.g., arithmetic, higher-order mathematics) assessed using different instruments. Through the use of a large sample and different competence measures, we can investigate whether individual differences in mathematical competencies are related to a few brain regions of the classical model of mathematical cognition—as suggested by Dehaene et al. (2003)—or whether a more complex picture emerges. Finally, to avoid the above-described pitfalls of VBM we applied SBM to examine differences in brain surface structure (i.e., cortical thickness, cortical surface complexity, gyrification index, and sulcal depth).

MATERIALS AND METHODS

Participants

Four-hundred and twenty-five German-speaking adults (female: $N = 271$; age: $M = 23.13$, $SD = 5.58$) participated in a larger behavioral test session, in which several psychometric measures were collected. Participants were recruited *via* social media, emails, and flyers. The majority of participants were enrolled as students at the University of Graz. All participants gave written informed consent before participation and received feedback regarding their intellectual abilities after testing as an incentive for taking part in the study.

From this pool, participants were recruited for two subsequent fMRI studies, in which, among other data, neurostructural images were collected. Both neuroimaging studies were carried out independently from each other. For the first neuroimaging study participants were randomly selected whereas in the second neuroimaging study participants were selected based on their arithmetic competencies (i.e., participants with lower and higher arithmetic competencies), which were measured in the behavioral test session. For both studies, all participants gave written informed consent before participation and were compensated with a minimum of € 20. The experimental procedure of both studies was approved by the ethics committee at the University of Graz, Austria.

For the current study, only participants from whom both data from the behavioral test session and neurostructural data were collected were included. In total 101 full data sets (Study 1: $N = 46$; Study 2: $N = 55$) were collected. However, due to partially missing data ($N = 10$) and other exclusion criteria (i.e., psychiatric disorders and left-handedness; $N = 2$) the final sample size comprises of 89 healthy young adults (Study 1: $N = 36$; Study 2: $N = 53$; age: $M = 21.88$; $SD = 3.43$; female: $N = 57$).

Materials and Stimuli of the Behavioral Test Session

Berlin Intelligence Structure Test (BIS-T)

Participants' numerical, verbal and figural intelligence was measured by the short version of the Berlin Intelligence Structure Test (Berliner Intelligenzstruktur-Test—BIS-T; Jäger et al., 1997). This test comprised 16 different tasks that either draw on the numerical (number of tests: 5), the verbal (number of tests: 6) or on the figural (number of tests: 5) component of intelligence. Additionally, each subtask can be assigned to four operational abilities, namely processing speed, memory, reasoning, and creativity. To assess numerical intelligence participants had to: (1) continue number series (reasoning); (2) cross out numbers in a matrix that were bigger by the factor of three in comparison to the preceding number (processing speed); (3) memorize pairs of digits (memory);

(4) estimate the results of complex calculations (reasoning); and (5) to find different operands resulting in a given arithmetic solution (creativity). Tasks assessing verbal intelligence required participants to (6) fill in a missing letter resulting in a given word (processing speed); (7) list as many abilities a good salesman should not have (creativity); (8) complete analogies (reasoning); (9) decide whether a given statement represents a fact or an opinion (reasoning); (10) answer questions to a memorized text (memory); and (11) to cross out words that are meronyms of the preceding word (e.g., "year" preceded by the word "month"; processing speed). Finally, to assess figural intelligence participants had to (12) complete figural analogies (reasoning); (13) memorize marked buildings on a city map (memory); (14) design as many logos for a bike shop (creativity); (15) cross out "x" in an array of letters (processing speed); and (16) to complete figures (reasoning). For the subsequent analysis, we used sum scores for each intelligence component (potential maximum scores: numerical = 100; verbal = 128; figural = 181) that were based on raw scores (i.e., number of correct answers) of each subtask.

Arithmetic Fluency Task

Arithmetic competencies were assessed through a paper-pencil task developed by Vogel et al. (2017b) that is based on the French Kit Test (French et al., 1963). This test measures performance on simple and complex arithmetic problems. Participants were presented with 128 simple multiplications (i.e., consisting of two single-digit operands), followed by 64 simple additions (i.e., consisting of two single-digit operands), 128 simple subtractions (i.e., consisting of an operand <20 and a single-digit operand), 60 complex multiplications (i.e., consisting of a double-digit and a single-digit operand), 60 complex additions (i.e., consisting of three double-digit operands) and, finally, 60 complex subtractions (i.e., consisting of two double-digit operands). Participants were instructed to solve as many problems as possible within a given time. For solving simple arithmetic problems participants were given 90 s for each operation and 120 s for each operation type of the complex arithmetic problems. We then computed two scores indicating the numbers of correctly solved problems separately for each complexity level (i.e., simple vs. complex), where 320 was the potential maximum score for simple arithmetic problems and 180 for complex arithmetic problems.

Mathematics Test for Selection of Personnel (M-PA)

We assessed the subject's performance in higher-order mathematics by using the short-version of the mathematics test for personnel selection (Mathematiktest für die Personalauswahl—M-PA; Jasper and Wagener, 2011). This test was developed to guarantee an optimal selection of suitable applicants based on their mathematical competencies. It is designed for teenagers and adults between 16 and 40 years who have at least a secondary school degree. The short version of the test comprises 31 complex mathematical problems presented as multiple-choice (MC) or open answer (OA) questions. Problems include fractions (3 OA), conversion of units (3 OA), exponentiation (7 OA), division with decimals (2 OA), algebra (1 MC), geometry (1 MC), roots (2 OA), and logarithm (7 OA). In total, participants had 15 min to solve the problems. For the subsequent analyses numbers of correctly solved problems were calculated (potential maximum score = 31).

Experimental Procedure

Behavioral data were collected in test sessions before the neuroimaging test sessions. Behavioral testing took place in laboratories of the Institute of Psychology at the University of Graz. One session took about 3 h, and participants were tested in small groups of a maximum of 12. After the greeting and general instruction participants were seated at a desk with a booklet, a computer screen and a keyboard on it. Desks were separated by partition walls, so that participants were able to work undisturbed. The booklet included a general instruction and all paper-pencil tests described in the "Materials and Methods" section. Additionally, the booklet included tests for measuring specific personality traits (i.e., the "NEO-Fünf-Faktoren Inventar—NEO-FFI" Borkenau and Ostendorf, 1993), the "German adaptation of the Abbreviated Math Anxiety Scale—AMAS-G" (Schillinger et al., 2018), the "Single-Item-Math Anxiety Scale—SIMA" (Núñez-Peña et al., 2014), a test measuring the attitude towards mathematics (Núñez-Peña et al., 2013, 2014), the "German Test Anxiety Inventory" (Prüfungsangstfragebogen, PAF; Hodapp et al., 2011), the "State-Trait Anxiety Inventory Trait scale—STAI-T" (Laux et al., 1981; Spielberger et al., 1983), and instructions for two computerized tasks (i.e., a single-digit multiplication task and an associative memory task), which participants also had to perform. The results of the latter tasks were not within the scope of the present study.

The order of tests was as following: BIS-T, M-PA, arithmetic fluency, computerized single-digit multiplication task, computerized associative memory task, NEO-FFI, PAF, attitudes towards mathematics, AMAS-G, SIMA, and STAI-T. Participants were instructed to work through the booklet page by page and to pause whenever they reached a page with a red stop sign. A trained experimenter took the time for all speeded tests (i.e., BIS-T: 45 min, M-PA: 15 min, arithmetic fluency: 10.5 min), and informed participants when they had to stop working on the respective test. At the end of the test booklet, participants were asked to fill out demographic questions according to their sex, age, field of study, and the mathematics grade in their final high school exam. In Austria, the grading system consists of a five-point scale, where 1 ("Sehr gut" – "very good") is the best possible grade and 5 ("Nicht genügend" – "unsatisfactory") is the lowest possible grade.

MRI Data Acquisition

Neuroimaging data were collected with a 3.0 T Siemens Skyra MRI scanner at the MRI Lab Graz. A 32-channel head coil and a Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA) sequence (TR = 1,950 ms, TE = 2.89 ms, $1 \times 1 \times 1$ mm isotropic voxel resolution) was used to acquire the high-resolution T1-weighted anatomical images.

In addition to high-resolution T1-weighted anatomical images other neuroimaging data were collected in both studies: In study 1, functional and diffusion MRI (fMRI and dMRI) data were acquired. Participants had to perform three different tasks

to measure the neural correlates of various cognitive processes. Results of the neurofunctional investigation are reported in Heidekum et al. (2019). In study 2 high-resolution T1-weighted anatomical images of participants were acquired within a training study that aimed to investigate the learning of arithmetic facts (i.e., multiplications). Participants underwent a 5-day multiplication fact training in which they intensively practiced a set of multiplications. Following the multiplication facts training, an MRI test session was conducted. In addition to the collection of anatomical data, functional imaging (i.e., task-based and resting-state activation data) as well as dMRI data were collected in this session. Since the behavioral and neurofunctional results of the training study are not within the scope of the present study they will be reported separately.

Analysis of Behavioral Data

Correlation coefficients were calculated to examine associations between all behavioral measurements. We calculated Spearman's correlation coefficients for investigating associations with mathematics grade, whilst for the rest of the associations, Pearson's correlation coefficients were calculated. *P*-values were Bonferroni corrected for 21 tested correlations (adjusted level of significance: $p_{Bonf} < 0.00238$). The normal distribution of each behavioral variable was verified by checking quantile-quantile plots and by calculating skewness and kurtosis. None of the variables significantly deviated from normality (i.e., $z_{Skewness} > 1.96$ or $z_{Kurtosis} > 1.96$).

Surface-Based Morphometry Analysis

SBM analyses were performed with the Computational Anatomy Toolbox (CAT12; Gaser and Dahnke, 2016), which is based on the Statistical Parametric Mapping (SPM, Welcome Department of Imaging Neuroscience, London, UK) software. In the first preprocessing step, T1-weighted anatomical images were normalized using "Diffeomorphic Anatomical Registration using Exponentiated Lie algebra" (DARTEL; Ashburner, 2007) and further segmented into GM, white matter (WM) and cerebrospinal fluid (CSF). In a next step, indices for cortical thickness (Dahnke et al., 2013), for cortical surface complexity (Yotter et al., 2011), for gyrification (Luders et al., 2006) and sulcal depth (Yotter et al., 2011) were calculated.

Cortical Thickness

Describes the distance between the inner surface (or boundary between GM and WM) and the outer surface (or boundary between GM and CSF; Dahnke et al., 2013). Cortical thickness, for instance, is an important biomarker for typical (O'Donnell et al., 2005) as well as atypical development (Thompson et al., 2004).

Fractal Dimensionality Index (Cortical Surface Complexity)

Characterizes the surface shape of the brain by quantifying the spatial frequency of gyrification and fissuration of the brain surface (Luders et al., 2004). More precisely, brain regions with a higher cortical surface complexity have a higher convolution level and therefore a larger surface area (Li et al., 2007; Yotter et al., 2011). Besides the investigation of age-related differences in brain structure (Madan and Kensinger, 2016), cortical complexity has also been successfully used to study differences in cognitive functions (King et al., 2010; Sandu et al., 2014).

Gyrification Index

Measures the regional curvature (i.e., convexity and concavity; Drury and Van Essen, 1997) of the brain and is defined as the ratio of the inner surface size to the outer surface size of an outer (usually convex) hull (Thompson et al., 1996; Luders et al., 2004, 2006). The gyrification index has typically been used to study group differences, such as between men and women (Luders et al., 2004) or between patients and controls (White et al., 2003; Shaw et al., 2012).

Sulcal Depth

Is based on the squared-transformed Euclidean distance between the central surface (average of the inner surface and the outer surface) and its convex hull (Lohmann, 1998; Yotter et al., 2011). Neuroanatomical studies using sulcal depth indices have shown, for instance, that it can account for differences in intelligence (Im et al., 2011; Yang et al., 2013).

Finally, full-width-at-half-maximum (FWHM) Gaussian kernels were used to smooth T1-weighted anatomical images. Following the matched-filter theorem, a 15.0 mm FWHM Gaussian kernel was applied for cortical thickness and a 20.0 mm one was used for folding measures (i.e., sulcal depth, gyrification, and cortical surface complexity).

In SPM we then performed second-level analyses to investigate whether structural variations in brain surface measures relate to individual differences in adults' mathematical competencies. Multiple regression analyses for each mathematical competence measure (i.e., mathematics grade, simple arithmetic, complex arithmetic, higher-order mathematics, and numerical intelligence) were calculated separately. The multiple regression analyses were performed for each brain surface measurement (i.e., cortical thickness, cortical surface complexity, gyrification, and sulcal depth) and were controlled for the influence of verbal and figural intelligence as well as for age and sex (i.e., additional regressors of no interest). Statistical results of these whole-brain analyses are reported with family-wise error (FWE) corrected values at the cluster level ($p < 0.05$).

RESULTS

Descriptive Analyses

In **Table 1** means, standard deviations and ranges for all behavioral measurements are displayed.

Correlation Analyses

Table 2 displays the correlations between all behavioral measurements. Almost all experimental measures that were thought to measure the same underlying component (i.e., mathematical cognition)—i.e., mathematics grade, simple arithmetic, complex arithmetic, higher-order mathematics and numerical intelligence—were significantly correlated with each other. The strongest association was found between

TABLE 1 | Descriptive statistics of all behavioral measurements (raw values).

	N	*M*	*SD*	Range
Mathematics grade[1]	89	2.47	0.96	(1–5)
Simple arithmetic	89	152.74	47.56	(52–246)
Complex arithmetic	89	47.63	19.28	(13–92)
Higher-order mathematics	89	21.20	5.49	(10–31)
Numerical intelligence	89	44.45	15.03	(12–87)
Verbal intelligence	89	92.11	16.14	(42–126)
Figural intelligence	89	84.12	11.97	(54–112)

Note: [1]Characteristics of the Austrian grading system: 1.00 = Very Good, 2.00 = Good, 3.00 = Satisfactory, 4 = Adequate, 5 = Unsatisfactory; Abbreviations: M = Mean; SD = Standard Deviation.

performance on simple and complex arithmetic problems ($r = 0.810$, $p < 0.001$).

For the mathematics grade, we found negative correlations with performance on complex arithmetic problems ($r = -0.358$, $p < 0.001$) as well as knowledge in higher-order mathematics ($r = -0.356$, $p < 0.001$) and numerical intelligence ($r = -0.346$, $p < 0.001$). This means that individuals with a better mathematics grade (associated with a smaller number) had better scores on experimental tests measuring mathematical abilities.

Additionally, both simple and complex arithmetic was positively correlated with verbal and figural intelligence. However, there was no significant correlation of mathematics grade with verbal and figural intelligence and no significant correlation of higher-order mathematics with verbal and figural intelligence. Finally, as can be expected, all three intelligence components (i.e., numerical, verbal and figural intelligence) were positively correlated with each other.

Structural Correlates of Mathematical Competencies

Multiple linear regression analyses were performed to investigate whether structural variations in brain surface measures (i.e., cortical thickness, cortical surface complexity, gyrification, and sulcal depth) relate to individual differences in adults' mathematical abilities (i.e., mathematics grade, simple arithmetic, complex arithmetic, higher-order mathematics and numerical intelligence). The results are as followed:

Cortical thickness: no significant negative or positive associations were found.

Cortical surface complexity: analyses revealed that numerical intelligence was positively associated with the cortical surface complexity of the right superior temporal gyrus [MNI (x, y, z): 60, −9, −7; $p < 0.05$ FWE cluster corrected; $k = 287$; **Figure 1A**]. No significant negative associations were found.

Gyrification: no significant negative or positive associations were found.

Sulcal depth: results showed a positive association between performance on complex arithmetic problems and the depth of the right central sulcus [MNI (x, y, z): 54, −9, 23; $p < 0.05$ FWE cluster corrected; $k = 141$; **Figure 1B**]. Additionally, multiple linear regression analysis revealed a positive relationship between adults' competence in higher-order mathematics and the depth of the left parieto-occipital sulcus [MNI (x, y, z): −18, −70, 29; $p < 0.05$ FWE cluster corrected; $k = 161$; **Figure 1C**]. No significant negative associations were found.

DISCUSSION

The current study is the first to investigate the associations between variations in brain surface structure (i.e., cortical thickness, cortical surface complexity, gyrification index, and sulcal depth) and individual differences in a broad range of mathematical abilities (i.e., mathematics grade, performance on simple and complex arithmetic problems, higher-order mathematics and numerical intelligence) within a large sample of typically developed adults. Analyses revealed three brain regions that were associated with individual differences in mathematical abilities: (1) the cortical surface complexity of the right superior temporal gyrus was positively related to numerical intelligence; (2) the depth of the right central sulcus was positively associated with individual's ability to solve complex arithmetic problems; and (3) the depth of the left parieto-occipital sulcus showed a positive relationship with the individual difference in performance on higher-order mathematics. Although SBM, as a tool to assess local brain morphology, does not allow any direct assumption of local brain function, this study supports the idea that these brain regions play a role in mathematical thinking. Surprisingly, we did not find any associations between adults' mathematical abilities and the neuroanatomical differences of previously reported brain regions, such as the IPS (Li et al., 2013; Lubin et al., 2013; Price et al., 2016). Conversely, all three brain regions that were observed in the present study are not typically associated with mathematics-related cognitive abilities. Despite these unexpected findings, the following sections will be an attempt to link functional properties to the observed cortical regions. Since the present study can only speculate on the functional role of these

TABLE 2 | Correlation coefficients between the behavioral measurements.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Mathematics grade	-						
(2) Simple arithmetic	−0.282	-					
(3) Complex arithmetic	−0.358*	0.810*	-				
(4) Higher-order mathematics	−0.356*	0.388*	0.471*	-			
(5) Numerical intelligence	−0.346*	0.681*	0.523*	0.379*	-		
(6) Verbal intelligence	−0.175	0.428*	0.454*	0.139	0.461*	-	
(7) Figural intelligence	−0.183	0.343*	0.325*	0.165	0.420*	0.370*	-

*$p < 0.00238$ (Bonferroni corrected p-value).

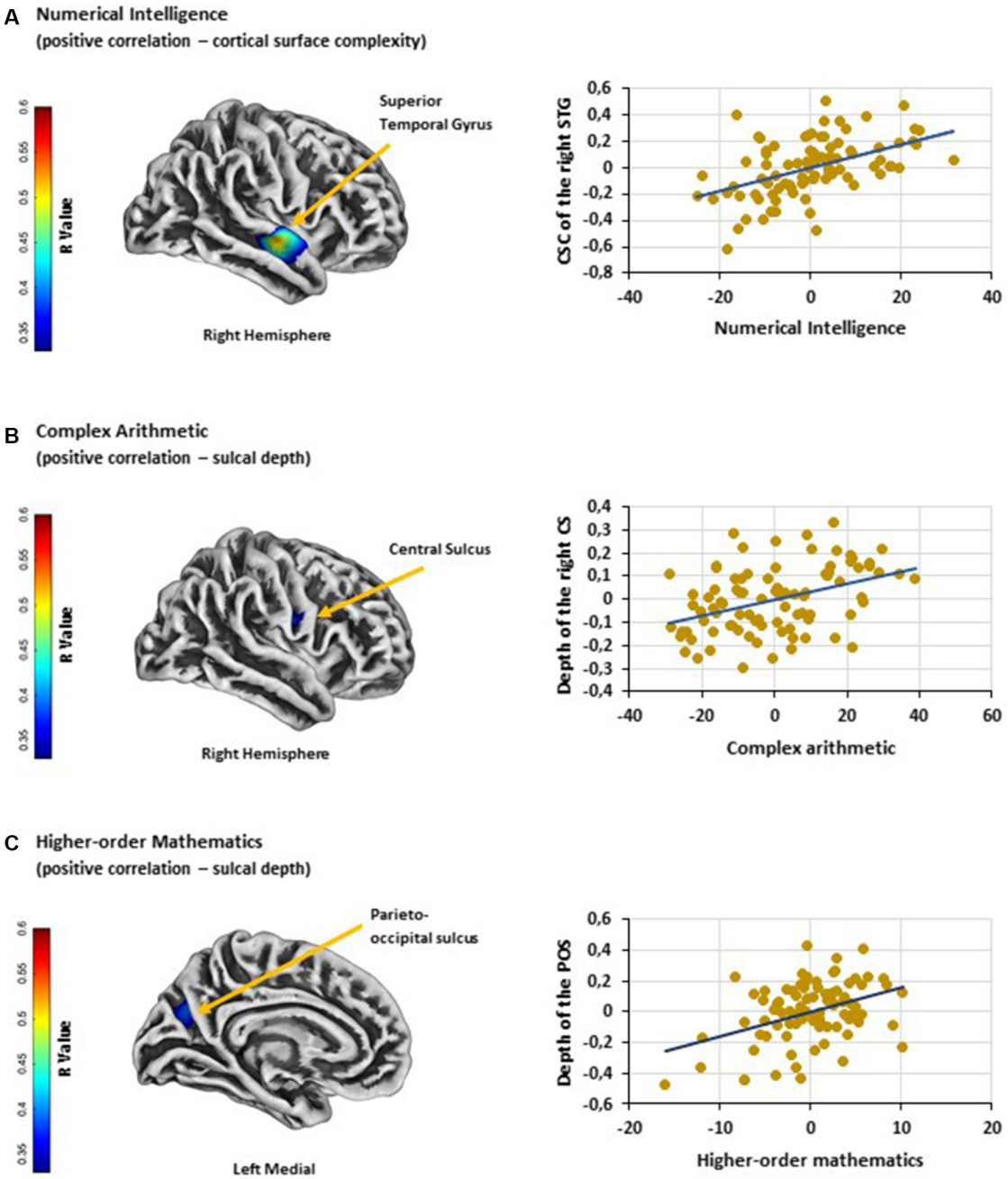


FIGURE 1 | (A) Significant positive association between numerical intelligence and cortical surface complexity of the right STG. **(B)** Significant positive association between performance on complex arithmetic problems and the depth of the right CS. **(C)** Significant positive association between higher-order mathematics competence and the depth of the left POS. Scatter plots showing partial correlations between **(A)** numerical intelligence and cortical surface complexity of the right STG, **(B)** performance on complex arithmetic problems and the depth of the right CS and **(C)** higher-order mathematics competence and the depth of the left POS. Colorbar: correlation coefficients (R-value). Abbreviations: STG, superior temporal gyrus; CSC, cortical surface complexity; CS, central sulcus; POS, parieto-occipital sulcus.

regions, further research (applying SBM for investigating brain-behavior relationships in the field of mathematical cognition) is needed.

First, our analyses revealed a positive relationship between the cortical surface complexity of the right superior temporal gyrus [MNI (x, y, z): 60, −9, −7] and individual differences in numerical intelligence. Surface complexity was defined by the fractal dimensionality index, which quantifies the spatial frequency of gyrification and fissuration of the brain surface (Luders et al., 2004). Thus, the current finding indicates that

in individuals with higher numerical intelligence the right superior temporal gyrus shows a higher convolution level and therefore a larger surface area. Cortical complexity has been successfully used to investigate age-related differences in brain structure (Madan and Kensinger, 2016) and, more importantly, to study differences in cognitive functions (King et al., 2010; Sandu et al., 2014). Unfortunately, the scarce neuroanatomical studies on the correlates of the right superior temporal gyrus are typically based on comparisons between various groups of patients and controls (e.g., schizophrenia: Honea et al., 2008; Nenadic et al., 2010; depression: Mak et al., 2009; Peng et al., 2011). Due to the methodological differences of these studies, they do not allow an interpretation of the current findings. Similarly, the results of neurofunctional studies are very heterogeneous. The right superior temporal gyrus has been implicated in several cognitive processes, such as spatial awareness (Karnath, 2001), emotion processing (Narumoto et al., 2001) or the detection of biological motion (Akiyama et al., 2006). Interestingly, previous work (Jung-Beeman et al., 2004) has also linked activity in the right superior temporal gyrus to creative problem solving, as it might be responsible for binding and accessing various types of available conceptual representations (Shen et al., 2017); cognitive processes that might also be important for problems in a numerical intelligence test. Besides these findings on domain-general cognitive functions, some studies have found activity in this region to be related to mathematical cognition (Fehr et al., 2008; Simos et al., 2008; Andres et al., 2011, 2012; Kucian et al., 2011; Gullick et al., 2012). Since the numerical intelligence test used in the current study required the correct application of arithmetic operations (i.e., continuing number series; crossing out numbers in a matrix that were bigger by the factor of three in comparison to the preceding number; finding different operands resulting in a given arithmetic solution), it is interesting that the majority of these studies (Fehr et al., 2008; Simos et al., 2008; Andres et al., 2011, 2012) has linked activation within this region to arithmetic problem-solving. For instance, in an fMRI block design, Andres et al. (2011) instructed a group of young males to multiply single Arabic digits (between 3 and 9) with, or subtract them from, a predefined digit. By contrasting multiplication problems with subtractions, the authors observed increased activity in right superior temporal areas (among other brain regions), which they related to the storing of semantic knowledge of arithmetic problems. Additionally, Fehr et al. (2008) investigated fMRI activation patterns during mental addition, subtraction, multiplication, and division in young adults. In contrast to Andres et al. (2011), greater right-hemispheric superior temporal activation was found by contrasting complex with simple division, which might have reflected the usage of procedural strategies for problem-solving. Accordingly, even though these findings are based on neurofunctional studies and they relate activity within the right superior temporal gyrus to different arithmetic strategies, they are in line with the result of the current study. Nevertheless, additional research is needed to confirm and to further specify the observed brain-behavior relationship.

Second, our results also revealed a positive correlation between the depth of the central sulcus [MNI (x, y, z): 54, −9, 23] and adults' ability to solve complex arithmetic problems. Previously, sulcal depth indices (e.g., sulcal fundi, lines or pits) have been successfully implemented to study the neuroanatomical correlates of various cognitive mechanisms. For instance, variations in the properties of sulci can account for individual differences in intelligence (Im et al., 2011; Yang et al., 2013), and abnormalities in specific sulci, such as the IPS or the superior temporal sulcus, have been observed in different groups of atypically developed individuals (e.g., dyslexia: Steinbrink et al., 2008; Richlan et al., 2013; dyscalculia: Rotzer et al., 2008; Price et al., 2016). Morphological changes of the central sulcus, on the other hand, have been associated with neurodegenerative diseases, such as multiple sclerosis (Pagani et al., 2005; Ceccarelli et al., 2008). However, the central sulcus also acts as an extension of its adjacent areas, namely the pre- and postcentral gyrus. Whereas the precentral gyrus is involved in eye movements (Corbetta et al., 1998; Anderson et al., 2007), postcentral activations have been linked to grasping (Simon et al., 2002)—both processes are commonly involved in cognitive tasks that include visually presented stimuli. For that reason, it is not surprising that these brain areas have also been found to play a role in numerical and calculation tasks (for a meta-analysis see Arsalidou and Taylor, 2011). For instance, Kaufmann et al. (2008) showed that activation patterns within these regions were related to non-symbolic numerical as well as spatial processing. And conversely, Kesler et al. (2006) have linked pre- and postcentral activations to the use of arithmetic strategies such as subvocalization and finger counting. In the present study, we observed a relationship between performance on complex arithmetic problems and the depth of the central sulcus. As the central sulcus acts as an extension of the pre- and postcentral gyrus, we, therefore, conclude that it might affect its adjacent mathematic-related regions.

Finally, the results of our SBM analyses revealed a positive association between the depth of the parieto-occipital sulcus [MNI (x, y, z): −18, −70, 29] and adult's higher-order mathematical competence. The parieto-occipital sulcus lies between the posterior parietal and occipital cortices. This brain region is a crucial node of the dorsal visual stream and is, therefore, involved in visuospatial processing that support spatial navigation and goal-directed actions (Caminiti et al., 1999; Hutchison et al., 2015; Tosoni et al., 2015; Richter et al., 2019). However, the parieto-occipital sulcus also separates the cuneus from the precuneus and lies in the direct vicinity of the IPS. Whereas the cuneus is associated with visual information processing (Vanni et al., 2001), the latter brain regions (i.e., precuneus and IPS) were found to contribute to higher neurocognitive functions, such as attention (Corbetta and Shulman, 2002; Rosen et al., 2015), working memory (Barton and Brewer, 2013; Bray et al., 2015) or memory retrieval (Hebscher et al., 2019; Heidekum et al., 2019). Moreover, the precuneus and the IPS are often found to be involved in mathematical cognition (Dehaene et al., 2003; Arsalidou and Taylor, 2011). In particular, previous studies in the field

of mathematical cognition have related brain activity in the precuneus to mental calculation (i.e., addition, multiplication, and subtraction; for a recent meta-analysis see Arsalidou and Taylor, 2011) and IPS activation to the representation and manipulation of numerical magnitude (Dehaene et al., 2003; or for a recent meta-analysis see Sokolowski et al., 2017). For instance, greater activation in the precuneus is often observed when arithmetic problems are solved by procedural strategies, rather than by memory retrieval (Delazer et al., 2005; Fehr et al., 2008). And, IPS activity is typically found whenever participants explicitly or implicitly compare the magnitude of numerals or dot-arrays (Pinel et al., 2001; Ansari et al., 2005; Vogel et al., 2017a; Wilkey et al., 2017). Interestingly, the functional role of the IPS within mathematics has also been confirmed by neurostructural studies (Isaacs et al., 2001; Rotzer et al., 2008; Rykhlevskaia et al., 2009; Ranpura et al., 2013) showing reduced GM in children with dyscalculia. In general, these studies highlight the importance of parietal brain regions in mathematical thinking, which is in line with the result of the current study showing an association between the parieto-occipital sulcus and higher-order mathematics.

Taken together, the current study highlights the potential of SBM as an alternative to classical VBM for investigating brain-behavior relationships in the domain of mathematical cognition. Associations were found between: (1) the cortical surface complexity of the right superior temporal gyrus and numerical intelligence; (2) the depth of the right central sulcus and adults' ability to solve complex arithmetic problems; and (3) the depth of the left parieto-occipital sulcus and adults' higher-order mathematics competence. Although two out of three brain regions (i.e., central sulcus and parieto-occipital sulcus) lie in intermediate vicinity of cortical structures typically associated with numerical and mathematical cognition, none of these brain regions were observed in previous studies. This discrepancy could be due to the methodological approach of the current study: In contrast to previous work, the present study studied linear brain-behavior relationships in a large sample of typically developed adults, whereas previous studies are mainly based on group comparisons between children with and without mathematical learning disabilities. For that reason, our results suggest that in typically developed individuals different neuroanatomical structures might become important when it comes to the processing of numerical and mathematical information. Finally, the current study shows that the depth of sulci could be a good index to investigate brain-behavior relationships. Nevertheless, future studies implementing SBM in different groups of individuals are needed.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics commitee of the University of Graz. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AH: conceptualization, data curation, formal analysis, writing—original draft. SV: conceptualization, formal analysis, funding acquisition, supervision, writing—original draft, writing—review and editing. RG: conceptualization, funding acquisition, project administration, supervision, writing—original draft, writing—review and editing.

FUNDING

This study was supported by a joint grant from the Research Foundation Flanders (Fonds Wetenschappelijk Onderzoek; FWO, G.0027.16) and the Austrian Science Fund (FWF, I 2425-G18).

ACKNOWLEDGMENTS

We thank Antonia Reuss, Dennis Wambacher and Thomas Zussner for assistance with data collection. Additionally, we thank Jochen Mosbacher for advice on the data analysis.

REFERENCES

Akiyama, T., Kato, M., Muramatsu, T., Saito, F., Nakachi, R., and Kashima, H. (2006). A deficit in discriminating gaze direction in a case with right superior temporal gyrus lesion. *Neuropsychologia* 44, 161–170. doi: 10.1016/j.neuropsychologia.2005.05.018

Anderson, E. J., Mannan, S. K., Husain, M., Rees, G., Sumner, P., Mort, D. J., et al. (2007). Involvement of prefrontal cortex in visual search. *Exp. Brain Res.* 180, 289–302. doi: 10.1007/s00221-007-0860-0

Andres, M., Michaux, N., and Pesenti, M. (2012). Common substrate for mental arithmetic and finger representation in the parietal cortex. *NeuroImage* 62, 1520–1528. doi: 10.1016/j.neuroimage.2012.05.047

Andres, M., Pelgrims, B., Michaux, N., Olivier, E., and Pesenti, M. (2011). Role of distinct parietal areas in arithmetic: an fMRI-guided TMS study. *NeuroImage* 54, 3048–3056. doi: 10.1016/j.neuroimage.2010.11.009

Ansari, D., Garcia, N., Lucas, E., Hamon, K., and Dhitalm, B. (2005). Neural correlates of symbolic number processing in children and adults. *Neuroreport* 16, 1769–1773. doi: 10.1097/01.wnr.0000183905.23396.f1

Arsalidou, M., and Taylor, M. J. (2011). Is 2+2=4? Meta-analyses of brain areas needed for numbers and calculations. *NeuroImage* 54, 2382–2393. doi: 10.1016/j.neuroimage.2010.10.009

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage* 38, 95–113. doi: 10.1016/j.neuroimage.2007.07.007

Aydin, K., Ucar, A., Oguz, K. K., Okur, O. O., Agayev, A., Unal, Z., et al. (2007). Increased gray matter density in the parietal cortex of mathematicians: a voxel-based morphometry study. *Am. J. Neuroradiol.* 28, 1859–1864. doi: 10.3174/ajnr.A0696

Barton, B., and Brewer, A. A. (2013). Visual working memory in human cortex. *Psychology* 4, 655–662. doi: 10.4236/psych.2013.48093

Berteletti, I., Prado, J. Ô., and Booth, J. R. (2014). Children with mathematical learning disability fail in recruiting verbal and numerical brain regions when solving simple multiplication problems. *Cortex* 57, 143–155. doi: 10.1016/j.cortex.2014.04.001

Borkenau, P., and Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI)*. Göttingen: Hogrefe.

Brans, R. G. H., Kahn, S., Schnack, H. G., Van Baal, G. C. M., Posthuma, D., Van Haren, N. E. M., et al. (2010). Brain plasticity and intellectual ability are influenced by shared genes. *J. Neurosci.* 30, 5519–5524. doi: 10.1523/JNEUROSCI.5841-09.2010

Bray, S., Almas, R., Arnold, A. E. G. F., Iaria, G., and MacQueen, G. (2015). Intraparietal sulcus activity and functional connectivity supporting spatial working memory manipulation. *Cereb. Cortex* 25, 1252–1264. doi: 10.1093/cercor/bht320

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376. doi: 10.1038/nrn3475

Caminiti, R., Genovesio, A., Marconi, B., Mayer, A. B., Onorati, P., Ferraina, S., et al. (1999). Early coding of reaching: frontal and parietal association connections of parieto-occipital cortex. *Eur. J. Neurosci.* 11, 3339–3345. doi: 10.1046/j.1460-9568.1999.00801.x

Cappelletti, M., and Price, C. J. (2014). Residual number processing in dyscalculia. *Neuroimage Clin.* 4, 18–28. doi: 10.1016/j.nicl.2013.10.004

Ceccarelli, A., Rocca, M. A., Pagani, E., Colombo, B., Martinelli, V., Comi, G., et al. (2008). A voxel-based morphometry study of grey matter loss in MS patients with different clinical phenotypes. *NeuroImage* 42, 315–322. doi: 10.1016/j.neuroimage.2008.04.173

Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. *R. Soc. Open Sci.* 1:140216. doi: 10.1098/rsos.140216

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., et al. (1998). A common network of functional areas for attention and eye movements. *Neuron* 21, 761–773. doi: 10.1016/s0896-6273(00)80593-0

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Dahnke, R., Yotter, R. A., and Gaser, C. (2013). Cortical thickness and central surface estimation. *NeuroImage* 65, 336–348. doi: 10.1016/j.neuroimage.2012.09.050

De Smedt, B., and Gilmore, C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *J. Exp. Child Psychol.* 108, 278–292. doi: 10.1016/j.jecp.2010.09.003

Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506. doi: 10.1080/02643290244000239

Delazer, M., Domahs, F., Bartha, L., Brenneis, C., Lochy, A., Trieb, T., et al. (2003). Learning complex arithmetic—an fMRI study. *Cogn. Brain Res.* 18, 76–88. doi: 10.1016/j.cogbrainres.2003.09.005

Delazer, M., Ischebeck, A., Domahs, F., Zamarian, L., Koppelstaetter, F., Siedentopf, C. M., et al. (2005). Learning by strategies and learning by drill—evidence from an fMRI study. *NeuroImage* 25, 838–849. doi: 10.1016/j.neuroimage.2004.12.009

Desai, R., Liebenthal, E., Possing, E. T., Waldron, E., and Binder, J. R. (2005). Volumetric vs. surface-based alignment for localization of auditory cortex activation. *NeuroImage* 26, 1019–1029. doi: 10.1016/j.neuroimage.2005.03.024

Drury, H. A., and Van Essen, D. C. (1997). Functional specializations in human cerebral cortex analyzed using the visible man surface-based atlas. *Hum. Brain Mapp.* 5, 233–237. doi: 10.1002/(sici)1097-0193(1997)5:4<233::aid-hbm5>3.0.co;2-4

Evans, T. M., Kochalka, J., Ngoon, T. J., Wu, S. S., Qin, S., Battista, C., et al. (2015). Brain structural integrity and intrinsic functional connectivity forecast 6 year longitudinal growth in children's numerical abilities. *J. Neurosci.* 35, 11743–11750. doi: 10.1523/JNEUROSCI.0216-15.2015

Fehr, T., Code, C., and Herrmann, M. (2008). Auditory task presentation reveals predominantly right hemispheric fMRI activation patterns during mental calculation. *Neurosci. Lett.* 431, 39–44. doi: 10.1016/j.neulet.2007.11.016

French, J. W., Ekstrom, R. B., and Price, L. A. (1963). *Manual for Kit of Reference Tests for Cognitive Factors (Revised 1963)*. Princeton, NJ: Educational Testing Service.

Gaser, C., and Dahnke, R. (2016). CAT – A computational anatomy toolbox for the analysis of structural MRI data. OHBM Conference 2016.

Geschwind, D. H., and Rakic, P. (2013). Cortical evolution: judge the brain by its cover. *Neuron* 80, 633–647. doi: 10.1016/j.neuron.2013.10.045

Grabner, R. H., Ansari, D., Reishofer, G., Stern, E., Ebner, F., and Neuper, C. (2007). Individual differences in mathematical competence predict parietal brain activation during mental calculation. *NeuroImage* 38, 346–356. doi: 10.1016/j.neuroimage.2007.07.041

Grabner, R. H., Ischebeck, A., Reishofer, G., Koschutnig, K., Delazer, M., Ebner, F., et al. (2009). Fact learning in complex arithmetic and figural-spatial tasks: the role of the angular gyrus and its relation to mathematical competence. *Hum. Brain Mapp.* 30, 2936–2952. doi: 10.1002/hbm.20720

Gullick, M. M., Wolford, G., and Temple, E. (2012). Understanding less than nothing: neural distance effects for negative numbers. *NeuroImage* 62, 542–554. doi: 10.1016/j.neuroimage.2012.04.058

Han, Z., Davis, N., Fuchs, L., Anderson, A. W., Gore, J. C., and Dawant, B. M. (2013). Relation between brain architecture and mathematical ability in children: a DBM study. *Magn. Reson. Imaging* 31, 1645–1656. doi: 10.1016/j.mri.2013.08.008

Han, Z., Fuchs, L., Davis, N., Cannistraci, C. J., Anderson, A. W., Gore, J. C., et al. (2008). "Analysis of anatomic variability in children with low mathematical skills," in *Proceedings of the Medical Imaging 2008: Physiology, Function and Structure from Medical Images*, San Diego, CA, 69160S.

Hebscher, M., Meltzer, J. A., and Gilboa, A. (2019). A causal role for the precuneus in network-wide theta and gamma oscillatory activity during complex memory retrieval. *Elife* 8:e43114. doi: 10.7554/eLife.43114

Heidekum, A. E., Grabner, R. H., De Smedt, B., De Visscher, A., and Vogel, S. E. (2019). Interference during the retrieval of arithmetic and lexico-semantic knowledge modulates similar brain regions: evidence from functional magnetic resonance imaging (fMRI). *Cortex* 120, 375–393. doi: 10.1016/j.cortex.2019.06.007

Henley, S. M. D., Ridgway, G. R., Scahill, R. I., Klöppel, S., Tabrizi, S. J., Fox, N. C., et al. (2010). Pitfalls in the use of voxel-based morphometry as a biomarker: examples from Huntington disease. *AJNR Am. J. Neuroradiol.* 31, 711–719. doi: 10.3174/ajnr.A1939

Hodapp, V., Rohrmann, S., and Ringeisen, T. (2011). *PAF—Prüfungsangstfragebogen*. Göttingen: Hogrefe.

Honea, R. A., Meyer-Lindenberg, A., Hobbs, K. B., Pezawas, L., Mattay, V. S., Egan, M. F., et al. (2008). Is gray matter volume an intermediate phenotype for schizophrenia? A voxel-based morphometry study of patients with schizophrenia and their healthy siblings. *Biol. Psychiatry* 63, 465–474. doi: 10.1016/j.biopsych.2007.05.027

Hutchison, R. M., Culham, J. C., Flanagan, J. R., Everling, S., and Gallivan, J. P. (2015). Functional subdivisions of medial parieto-occipital cortex in humans and nonhuman primates using resting-state fMRI. *NeuroImage* 116, 10–29. doi: 10.1016/j.neuroimage.2015.04.068

Hutton, C., Draganski, B., Ashburner, J., and Weiskopf, N. (2009). A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging. *NeuroImage* 48, 371–380. doi: 10.1016/j.neuroimage.2009.06.043

Im, K., Choi, Y. Y., Yang, J. J., Lee, K. H., Kim, S. I., Grant, P. E., et al. (2011). The relationship between the presence of sulcal pits and intelligence in human brains. *NeuroImage* 55, 1490–1496. doi: 10.1016/j.neuroimage.2010.12.080

Isaacs, E. B., Edmonds, C. J., Lucas, A., and Gadian, D. G. (2001). Calculation difficulties in children of very low birthweight: a neural correlate. *Brain* 124, 1701–1707. doi: 10.1093/brain/124.9.1701

Ischebeck, A., Zamarian, L., Schocke, M., and Delazer, M. (2009). Flexible transfer of knowledge in mental arithmetic—an fMRI study. *NeuroImage* 44, 1103–1112. doi: 10.1016/j.neuroimage.2008.10.025

Ischebeck, A., Zamarian, L., Siedentopf, C., Koppelstätter, F., Benke, T., Felber, S., et al. (2006). How specifically do we learn? Imaging the learning of multiplication and subtraction. *NeuroImage* 30, 1365–1375. doi: 10.1016/j.neuroimage.2005.11.016

Jäger, A. O., Süß, H.-M., and Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test [Berlin Intelligence Structure Test]*. Göttingen: Hogrefe.

Jasper, F., and Wagener, D. (2011). *Mathematiktest für die Personalauswahl [Mathematics Test for Selection of Personnel]*. Göttingen: Hogrefe.

Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., et al. (2004). Neural activity when people solve verbal problems with insight. *PLoS Biol.* 2:E97. doi: 10.1371/journal.pbio.0020097

Karnath, H. O. (2001). New insights into the functions of the superior temporal cortex. *Nat. Rev. Neurosci.* 2, 568–576. doi: 10.1038/35086057

Kaufmann, L., Vogel, S. E., Wood, G., Kremser, C., Schocke, M., Zimmerhackl, L. B., et al. (2008). A developmental fMRI study of nonsymbolic numerical and spatial processing. *Cortex* 44, 376–385. doi: 10.1016/j.cortex.2007.08.003

Kesler, S. R., Menon, V., and Reiss, A. L. (2006). Neurofunctional differences associated with arithmetic processing in Turner syndrome. *Cereb. Cortex* 16, 849–856. doi: 10.1093/cercor/bhj028

King, R. D., Brown, B., Hwang, M., Jeon, T., and George, A. T. (2010). Fractal dimension analysis of the cortical ribbon in mild Alzheimer's disease. *NeuroImage* 53, 471–479. doi: 10.1016/j.neuroimage.2010.06.050

Klein, E., Moeller, K., Glauche, V., Weiller, C., and Willmes, K. (2013). Processing pathways in mental arithmetic—evidence from probabilistic fiber tracking. *PLoS One* 8:e55455. doi: 10.1371/journal.pone.0055455

Klein, E., Suchan, J., Moeller, K., Karnath, H.-O. O., Knops, A., Wood, G., et al. (2014). Considering structural connectivity in the triple code model of numerical cognition: differential connectivity for magnitude processing and arithmetic facts. *Brain Struct. Funct.* 221, 979–995. doi: 10.1007/s00429-014-0951-1

Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönmann, C., Plangger, F., et al. (2011). Mental number line training in children with developmental dyscalculia. *NeuroImage* 57, 782–795. doi: 10.1016/j.neuroimage.2011.01.070

Laux, L., Glanzmann, P., Schaffner, P., and Spielberger, C. (1981). *Das State-Trait-Angstinventar: STAI*. Weinheim: Beltz.

Li, Y., Hu, Y., Wang, Y., Weng, J., and Chen, F. (2013). Individual structural differences in left inferior parietal area are associated with schoolchildrens' arithmetic scores. *Front. Hum. Neurosci.* 7:844. doi: 10.3389/fnhum.2013.00844

Li, X., Jiang, J., Zhu, W., Yu, C., Sui, M., Wang, Y., et al. (2007). Asymmetry of prefrontal cortical convolution complexity in males with attention-deficit/hyperactivity disorder using fractal information dimension. *Brain Dev.* 29, 649–655. doi: 10.1016/j.braindev.2007.04.008

Lohmann, G. (1998). Extracting line representations of sulcal and gyral patterns in MR images of the human brain. *IEEE Trans. Med. Imaging* 17, 1040–1048. doi: 10.1109/42.746714

Lubin, A., Rossi, S., Simon, G., Lanoë, C., Leroux, G., Poirel, N., et al. (2013). Numerical transcoding proficiency in 10-year-old schoolchildren is associated with gray matter inter-individual differences: a voxel-based morphometry study. *Front. Psychol.* 4:197. doi: 10.3389/fpsyg.2013.00197

Luders, E., Narr, K. L., Thompson, P. M., Rex, D. E., Jancke, L., Steinmetz, H., et al. (2004). Gender differences in cortical complexity. *Nat. Neurosci.* 7, 799–800. doi: 10.1038/nn1277

Luders, E., Thompson, P. M., Narr, K. L., Toga, A. W., Jancke, L., and Gaser, C. (2006). A curvature-based approach to estimate local gyrification on the cortical surface. *NeuroImage* 29, 1224–1230. doi: 10.1016/j.neuroimage.2005.08.049

Madan, C. R., and Kensinger, E. A. (2016). Cortical complexity as a measure of age-related brain atrophy. *NeuroImage* 134, 617–629. doi: 10.1016/j.neuroimage.2016.04.029

Mak, A. K. Y., Wong, M. M. C., Han, S. H., and Lee, T. M. C. (2009). Gray matter reduction associated with emotion regulation in female outpatients with major depressive disorder: a voxel-based morphometry study. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 33, 1184–1190. doi: 10.1016/j.pnpbp.2009.06.025

Menon, V. (2014). "Arithmetic in the child and adult brain," in *The Oxford Handbook of Mathematical Cognition*, eds R. Cohen Kadosh and A. Dowker (Oxford: Oxford University Press), 1–23.

Moreau, D., Wiebels, K., Wilson, A. J., and Waldie, K. E. (2019). Volumetric and surface characteristics of gray matter in adult dyslexia and dyscalculia. *Neuropsychologia* 127, 204–210. doi: 10.1016/j.neuropsychologia.2019.02.002

Narumoto, J., Okada, T., Sadato, N., Fukui, K., and Yonekura, Y. (2001). Attention to emotion modulates fMRI activity in human right superior temporal sulcus. *Cogn. Brain Res.* 12, 225–231. doi: 10.1016/s0926-6410(01)00053-2

Nenadic, I., Smesny, S., Schlösser, R. G. M., Sauer, H., and Gaser, C. (2010). Auditory hallucinations and brain structure in schizophrenia: voxel-based morphometric study. *Br. J. Psychiatry* 196, 412–413. doi: 10.1192/bjp.bp.109.070441

Núñez-Peña, M. I., Guilera, G., and Suárez-Pellicioni, M. (2014). The single-item math anxiety scale: an alternative way of measuring mathematical anxiety. *J. Psychoeduc. Assess.* 32, 306–317. doi: 10.1177/0734282913508528

Núñez-Peña, M. I., Suárez-Pellicioni, M., Guilera, G., and Mercadé-Carranza, C. (2013). A spanish version of the short mathematics anxiety rating scale (sMARS). *Learn. Individ. Differ.* 24, 204–210. doi: 10.1016/j.lindif.2012.12.009

O'Donnell, S., Noseworthy, M. D., Levine, B., and Dennis, M. (2005). Cortical thickness of the frontopolar area in typically developing children and adolescents. *NeuroImage* 24, 948–954. doi: 10.1016/j.neuroimage.2004.10.014

Pagani, E., Rocca, M. A., Gallo, A., Rovaris, M., Martinelli, V., Comi, G., et al. (2005). Regional brain atrophy evolves differently in patients with multiple sclerosis according to clinical phenotype. *Am. J. Neuroradiol.* 26, 341–346.

Parsons, S., and Bynner, J. (2005). *Does Numeracy Matter More?* London: National Research and Development Centre for Adult Literacy and Numeracy.

Peelle, J. E., Cusack, R., and Henson, R. N. A. (2012). Adjusting for global effects in voxel-based morphometry: gray matter decline in normal aging. *NeuroImage* 60, 1503–1516. doi: 10.1016/j.neuroimage.2011.12.086

Peng, J., Liu, J., Nie, B., Li, Y., Shan, B., Wang, G., et al. (2011). Cerebral and cerebellar gray matter reduction in first-episode patients with major depressive disorder: a voxel-based morphometry study. *Eur. J. Radiol.* 80, 395–399. doi: 10.1016/j.ejrad.2010.04.006

Peters, L., and De Smedt, B. (2018). Arithmetic in the developing brain: a review of brain imaging studies. *Dev. Cogn. Neurosci.* 30, 265–279. doi: 10.1016/j.dcn.2017.05.002

Pinel, P., Dehaene, S., Rivière, D., and LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. *NeuroImage* 14, 1013–1026. doi: 10.1006/nimg.2001.0913

Price, G. R., Wilkey, E. D., Yeo, D. J., and Cutting, L. E. (2016). The relation between 1st grade grey matter volume and 2nd grade math competence. *NeuroImage* 124, 232–237. doi: 10.1016/j.neuroimage.2015.08.046

Ranpura, A., Isaacs, E., Edmonds, C., Rogers, M., Lanigan, J., Singhal, A., et al. (2013). Developmental trajectories of grey and white matter in dyscalculia. *Trends Neurosci. Educ.* 2, 56–64. doi: 10.1016/j.tine.2013.06.007

Richlan, F., Kronbichler, M., and Wimmer, H. (2013). Structural abnormalities in the dyslexic brain: a meta-analysis of voxel-based morphometry studies. *Hum. Brain Mapp.* 34, 3055–3065. doi: 10.1002/hbm.22127

Richter, M., Amunts, K., Mohlberg, H., Bludau, S., Eickhoff, S. B., Zilles, K., et al. (2019). Cytoarchitectonic segregation of human posterior intraparietal and adjacent parieto-occipital sulcus and its relation to visuomotor and cognitive functions. *Cereb. Cortex* 29, 1305–1327. doi: 10.1093/cercor/bhy245

Rosen, M. L., Stern, C. E., Michalka, S. W., Devaney, K. J., and Somers, D. C. (2015). Influences of long-term memory-guided attention and stimulus-guided attention on visuospatial representations within human intraparietal sulcus. *J. Neurosci.* 35, 11358–11363. doi: 10.1523/JNEUROSCI.1055-15.2015

Rotzer, S., Kucian, K., Martin, E., von Aster, M., Klaver, P., and Loenneker, T. (2008). Optimized voxel-based morphometry in children with developmental dyscalculia. *NeuroImage* 39, 417–422. doi: 10.1016/j.neuroimage.2007.08.045

Rykhlevskaia, E., Uddin, L. Q., Kondos, L., and Menon, V. (2009). Neuroanatomical correlates of developmental dyscalculia: combined evidence from morphometry and tractography. *Front. Hum. Neurosci.* 3:51. doi: 10.3389/neuro.09.051.2009

Sandu, A. L., Staff, R. T., McNeil, C. J., Mustafa, N., Ahearn, T., Whalley, L. J., et al. (2014). Structural brain complexity and cognitive decline in late life—a longitudinal study in the Aberdeen 1936 Birth Cohort. *NeuroImage* 100, 558–563. doi: 10.1016/j.neuroimage.2014.06.054

Schillinger, F. L., Vogel, S. E., Diedrich, J., and Grabner, R. H. (2018). Math anxiety, intelligence, and performance in mathematics: insights from the German adaptation of the Abbreviated Math Anxiety Scale (AMAS-G). *Learn. Individ. Differ.* 61, 109–119. doi: 10.1016/j.lindif.2017.11.014

Schnack, H. G., Van Haren, N. E. M., Brouwer, R. M., Evans, A., Durston, S., Boomsma, D. I., et al. (2015). Changes in thickness and surface area of the human cortex and their relationship with intelligence. *Cereb. Cortex* 25, 1608–1617. doi: 10.1093/cercor/bht357

Shaw, P., Malek, M., Watson, B., Sharp, W., Evans, A., and Greenstein, D. (2012). Development of cortical surface area and gyrification in attention-deficit/hyperactivity disorder. *Biol. Psychiatry* 72, 191–197. doi: 10.1016/j.biopsych.2012.01.031

Shen, W., Yuan, Y., Liu, C., and Luo, J. (2017). The roles of the temporal lobe in creative insight: an integrated review. *Think. Reason.* 23, 321–375. doi: 10.1080/13546783.2017.1308885

Simon, O., Mangin, J. F., Cohen, L., Le Bihan, D., and Dehaene, S. (2002). Topographical layout of hand, eye, calculation and language-related areas in the human parietal lobe. *Neuron* 33, 475–487. doi: 10.1016/s0896-6273(02)00575-5

Simos, P. G., Kanatsouli, K., Fletcher, J. M., Sarkari, S., Juranek, J., Cirino, P., et al. (2008). Aberrant spatiotemporal activation profiles associated with math difficulites in children: a magnetic source imaging study. *Neuropsychology* 22, 571–584. doi: 10.1037/0894-4105.22.5.571

Sokolowski, H. M., Fias, W., Mousa, A., and Ansari, D. (2017). Common and distinct brain regions in both parietal and frontal cortex support symbolic and nonsymbolic number processing in humans: a functional neuroimaging meta-analysis. *NeuroImage* 146, 376–394. doi: 10.1016/j.neuroimage.2016.10.028

Spielberger, C. D., Gorsuch, R. L., Lushene, P. R., Vagg, P. R., and Jacobs, A. G. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Starke, M., Kiechl-Kohlendorfer, U., Kucian, K., Pupp Peglow, U., Kremser, C., Schocke, M., et al. (2013). Brain structure, number magnitude processing and math proficiency in 6- to 7-year-old children born prematurely: a voxel-based morphometry study. *Neuroreport* 24, 419–424. doi: 10.1097/WNR.0b013e32836140ed

Steinbrink, C., Vogt, K., Kastrup, A., Müller, H. P., Juengling, F. D., Kassubek, J., et al. (2008). The contribution of white and gray matter differences to developmental dyslexia: insights from DTI and VBM at 3.0 T. *Neuropsychologia* 46, 3170–3178. doi: 10.1016/j.neuropsychologia.2008.07.015

Supekar, K., Swigart, A. G., Tenison, C., Jolles, D. D., Rosenberg-Lee, M., Fuchs, L., et al. (2013). Neural predictors of individual differences in response to math tutoring in primary-grade school children. *Proc. Natl. Acad. Sci. U S A* 110, 8230–8235. doi: 10.1073/pnas.1222154110

Thompson, P. M., Hayashi, K. M., Sowell, E. R., Gogtay, N., Giedd, J. N., Rapoport, J. L., et al. (2004). Mapping cortical change in Alzheimer's disease, brain development and schizophrenia. *NeuroImage* 23, 2–18. doi: 10.1016/j.neuroimage.2004.07.071

Thompson, P. M., Schwartz, C., Lin, R. T., Khan, A. A., and Toga, A. W. (1996). Three-dimensional statistical analysis of sulcal variability in the human brain. *J. Neurosci.* 16, 4261–4274. doi: 10.1523/jneurosci.16-13-04261.1996

Tosoni, A., Pitzalis, S., Committeri, G., Fattori, P., Galletti, C., and Galati, G. (2015). Resting-state connectivity and functional specialization in human medial parieto-occipital cortex. *Brain Struct. Funct.* 220, 3307–3321. doi: 10.1007/s00429-014-0858-x

Vanni, S., Tanskanen, T., Seppä, M., Uutela, K., and Hari, R. (2001). Coinciding early activation of the human primary visual cortex and anteromedial cuneus. *Proc. Natl. Acad. Sci. U S A* 98, 2776–2780. doi: 10.1073/pnas.041600898

Voets, N. L., Hough, M. G., Douaud, G., Matthews, P. M., James, A., Winmill, L., et al. (2008). Evidence for abnormalities of cortical development in adolescent-onset schizophrenia. *NeuroImage* 43, 665–675. doi: 10.1016/j.neuroimage.2008.08.013

Vogel, S. E., Goffin, C., Bohnenberger, J., Koschutnig, K., Reishofer, G., Grabner, R. H., et al. (2017a). The left intraparietal sulcus adapts to symbolic number in both the visual and auditory modalities: evidence from fMRI. *NeuroImage* 153, 16–27. doi: 10.1016/j.neuroimage.2017.03.048

Vogel, S. E., Haigh, T., Sommerauer, G., Spindler, M., Brunner, C., Lyons, I. M., et al. (2017b). Processing the order of symbolic numbers: a reliable and unique predictor of arithmetic fluency. *J. Numer. Cogn.* 3, 288–308. doi: 10.5964/jnc.v3i2.55

White, T., Andreasen, N. C., Nopoulos, P., and Magnotta, V. (2003). Gyrification abnormalities in childhood- and adolescent-onset schizophrenia. *Biol. Psychiatry* 54, 418–426. doi: 10.1016/s0006-3223(03)00065-9

Wilkey, E. D., Barone, J. C., Mazzocco, M. M. M., Vogel, S. E., and Price, G. R. (2017). The effect of visual parameters on neural activation during nonsymbolic number comparison and its relation to math competency. *NeuroImage* 159, 430–442. doi: 10.1016/j.neuroimage.2017.08.023

Wilkey, E. D., Cutting, L. E., and Price, G. R. (2018). Neuroanatomical correlates of performance in a state-wide test of math achievement. *Dev. Sci.* 21:2. doi: 10.1111/desc.12545

Yang, J. J., Yoon, U., Yun, H. J., Im, K., Choi, Y. Y., Lee, K. H., et al. (2013). Prediction for human intelligence using morphometric characteristics of cortical surface: partial least square analysis. *Neuroscience* 246, 351–361. doi: 10.1016/j.neuroscience.2013.04.051

Yotter, R. A., Nenadic, I., Ziegler, G., Thompson, P. M., and Gaser, C. (2011). Local cortical surface complexity maps from spherical harmonic reconstructions. *NeuroImage* 56, 961–973. doi: 10.1016/j.neuroimage.2011.02.007

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

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership