National Security and the Assessment of Individual Credibility: Current Challenges, Future Opportunities

Mark D. Happel, Jason A. Spitaletta, Eric A. Pohlmeyer, Grace M. Hwang, Ariel M. Greenberg, Clara A. Scholl, and Michael Wolmetz

ABSTRACT

This article examines the current technology-based capabilities of national security and law enforcement officials to assess the credibility of individuals who are being evaluated as a potential source of information or to determine whether they can be trusted with sensitive information. At present, these officials, both domestically and internationally, rely most heavily on the polygraph for a wide variety of credibility assessment applications. However, its accuracy and reliability vary greatly across the different investigative problems to which it is applied. Major improvements in credibility assessment will likely require considerable investments in basic research, but more modest improvements appear within reach by using existing instruments and methods. Perhaps the most promising is the electroencephalogram (EEG), which may be able to detect when an individual is attempting to conceal information. The applicable EEG-based credibility assessment research is reviewed, showing limited but realistic potential for near-term application to some credibility assessment applications.

INTRODUCTION

> The problem presented by many of the new threats, whether from transnational terrorist groups or from non-traditional nation-state adversaries, however, is not that of accessing denied areas but of penetrating "denied minds"—and not just those of a few recognized leaders, but of groups, social networks, and entire cultures.[1]

The ability to gain knowledge of an adversary's intentions and capabilities remains at the core of national intelligence priorities. However, in this era of asymmetric threats and violent non-state actors, the need to obtain credible knowledge directly from individuals is greater than ever. All people tell lies or conceal information from time to time,[2] and even trained security professionals are typically not much better than chance at detecting another person's deceptive responses.[3–5] Nevertheless, in many critical situations, government officials and law enforcement personnel must be able to distinguish between a credible source of key information and one who cannot be counted on to relate a full and accurate account.

Credibility assessment is the term used by the U.S. DoD for "instrumentation, techniques, and procedures to assess the truthfulness and credibility of individuals."[6] For the purpose of this article, credibility assessment will be interpreted as assessing the extent to which an individual does not *intentionally* attempt to deceive a government official in the performance of his/her official duties. Deception itself can be defined so that

SPECIAL FEATURE

social “white lies” are not included by restricting it to “the intentional concealment, distortion, or fabrication of information for the purpose of gaining an advantage or leading another into an erroneous conclusion.”[7] Credibility assessment then covers the evaluation of a given individual as a source of information, attempting to identify distorted or fabricated information and when critical information is being concealed. In a prospective sense, it also attempts to determine whether someone can be trusted with sensitive information, with no intention to do harm to national security or to others.

Current approaches to credibility assessment can be roughly divided into behavioral and technology-based approaches (sometimes referred to as mechanical approaches).[8] The polygraph examination serves as the prototypical example of a technology-based approach, while the cognitive interview, a method initiated within child clinical psychology and migrated to intelligence and law enforcement that seeks to improve the subject’s memory retrieval processes, is a good example of a behavioral approach.[9,10] The distinction between behavioral and technology-based approaches is useful, but the boundary is not very well defined. For example, the polygraph exam benefits from the use of a well-conducted behavioral interview before the polygraph instrument is used, and behavioral methods can benefit from videotaping and replaying for subsequent analysis. This article focuses on the technology-based approaches; further information regarding behavioral approaches can be found by consulting the applicable references.[9–13]

Two important distinctions can be made among the many credibility assessment applications: compliance and content. With compliant participants, the investigator has greater control over the situation and can more readily use the instrument as a “psychological anvil” to compel additional disclosures of information that is relevant to the process but may not have been revealed in a background investigation. However, those accused of a crime, or those who may possess information of national intelligence value, might not be as compliant or may use countermeasures to attempt to thwart the examination. Research to date has often relied on compliant participants, and the DoD considers the results of research into credibility assessment countermeasures to be classified. The content of an assessment can vary between an individual’s past experiences (retrospective behavior) and an individual’s future intentions (prospective behavior). Forensic criminal investigations focus on retrospective behavior, while initial security screenings often focus on prospective behavior. Notably, while there has been considerable psychological, neuroscientific, and forensic research on retrospective memory and report, very little research applies to prospective behavior examinations. Some important credibility assessment application areas, organized by assumed compliance and content, are shown in Fig. 1.

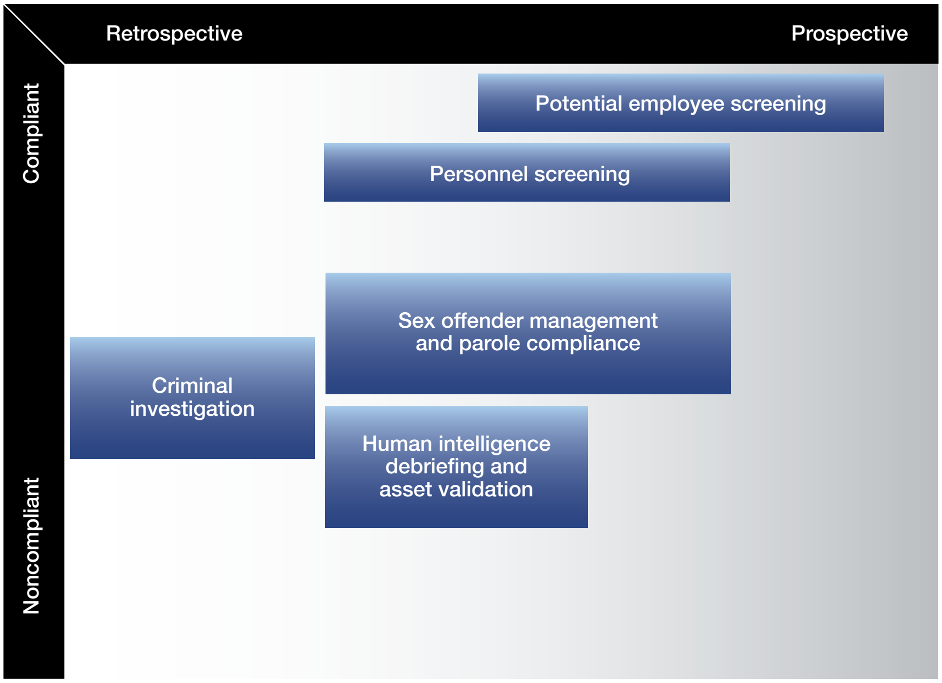
Retrospective
Prospective
Compliant
Noncompliant
Potential employee screening
Personnel screening
Sex offender management and parole compliance
Criminal investigation
Human intelligence debriefing and asset validation

Figure 1. Summary of application space in terms of compliance and content. Placement along the horizontal axis represents whether the content is more retrospective or prospective, and placement along the vertical axis is the degree of compliance expected (i.e., how much of a concern are countermeasures in terms of likelihood and sophistication). Note that these are highly variable applications, and this figure can convey only a general sense of the application space.

THE POLYGRAPH

The polygraph instrument is at present by far the most widely used means of technology-based credibility assessment. The polygraph is currently used by the DoD and the intelligence community for personnel screening, asset validation, and criminal investigations. Four organizations within the Department of Justice—the Federal Bureau of Investigation; the Drug Enforcement Administration; the Bureau of Alcohol, Tobacco, Firearms and Explosives; and the Department of Justice Office of the Inspector General—use the polygraph in fulfillment of law enforcement responsibilities (and, in the case of the Federal Bureau of Investigation, counterintelligence activities). Polygraphy is also an important tool for state and local law enforcement agencies, even though the results of a polygraph examination are inadmissible as evidence in court. Probation and parole agencies in over 35 U.S. states use the polygraph in the supervision of sex offenders. In adult community or institutionally based programs, over 50% and 80% of treatment programs incorporate the instrument, respectively.[14,15]

A typical modern polygraph instrument consists of sensors for recording physiological measures, appropriate amplifiers and signal conditioning hardware, and a laptop computer for signal processing, recording, and displaying the collected data. The most common set of physiological measurements recorded by the polygraph are blood pressure/heart rate, respiration rate (measured at both the abdomen and the chest), and skin conductance.[16] All four measures are displayed on the polygraph instrument as a function of time. In general, polygraph examiners are trained to look for anomalies in each of the measurement waveforms independently, although some available scoring algorithms seek to treat all of the measures simultaneously. Not all the measurements need to be used in each of the possible test methods.

The polygraph instrument itself, however, is capable of measuring only time courses of physiological responses, such as heart rate. To use these signals to infer intentional attempts to deceive the examiner, it is necessary to create a social and psychological situation in which the physiological signals can be properly interpreted. This structure within which the polygraph instrument can be employed is referred to as the polygraph test. The standard polygraph test can be viewed as a controlled psychophysiological experiment involving (*i*) a trained polygraph examiner who conducts the experiment, (*ii*) a test subject (or participant) who will reply (usually verbally) in response to questions and/or audiovisual stimuli, and (*iii*) an instrument capable of recording and displaying a set of measurements of the subject's autonomic responses. Federal polygraph tests usually also include a supervisor in a remote location who monitors the conduct of the exam via video and audio feeds and can observe the polygraph screen via a remote monitor.

During the test, the examiner asks a series of questions or presents stimuli (images, audio, etc.), and the subject generally responds with simple verbal responses (e.g., yes/no). The subject is instructed to relax, fix their gaze on some meaningless location (e.g., a spot on the wall), and refrain from extraneous muscle movement (which could confound the psychophysiological signals being measured and could be interpreted by the examiner as an attempt to defeat the machine). After the test has been completed, the polygrapher generally reviews any adverse findings with the subject, in the hope that any misunderstanding or procedural errors can be accounted for and corrected. If the subject was deemed to have been deceptive on one or more questions, he/she may be given the opportunity to provide additional clarifying information or to substitute more truthful responses. If the overall test is judged to be inconclusive, a follow-on examination may be scheduled for a later date.

Despite its pervasive use, a single unified policy for conducting and interpreting polygraph examinations within the United States (or even across all agencies of the federal government) does not exist, leading to a wide variety of procedures and standards for its employment. Nevertheless, polygraph tests can be grouped into three general classes (although there are significant variations within each class):

1. The **Relevant/Irrelevant Test (RIT)** compares the subject's responses to two different sets of questions: the irrelevant set of questions (e.g., confirming that his or her age is equal to a specific value) and the relevant set of questions. The test assumes that guilty subjects will react more strongly to the relevant questions than to the irrelevant questions.

2. The **Comparison Question Test (CQT)** also uses two sets of questions but replaces the neutral irrelevant set with a comparison question (or probable lie) set. The comparison set includes questions about unethical acts that would cause most innocent subjects to lie to protect their egos or reputations. The test assumes that guilty subjects will react to the relevant questions more than to the comparison questions.

3. The **Concealed Information Test (CIT)**, also known as the Guilty Knowledge Test (GKT), uses a single set of multiple-choice questions, each with five or more possible answers (or alternatively, presentations of items such as photographs), only one of which reflects the "true" state of affairs. The test assumes that subjects who are concealing information about the "true" item will react more strongly to that item compared to the accompanying fictitious ones.

These tests are summarized in Table 1.

It is important to note that the success or failure of the polygraph test is not solely dependent on the accuracy of the polygraph device alone; the interactions between the examiner and the subject are frequently more revealing than the specific traces displayed on the polygraph instrument. The polygrapher watches the subject carefully throughout the test (generally from a position in which the subject cannot also observe him/her) for behavioral indications of deception or signs that the subject is attempting to employ countermeasures. Although the deception detection skills the polygrapher uses contribute valuably to meeting the test's goals, they also complicate any attempt to objectively evaluate the accuracy of a given polygraph test: the result is not a function of the instrument alone but also of the observational and interpretive skills of the polygrapher.

Table 1. Principal classes of polygraph examinations

Test	RIT	CQT	CIT
General approach	Compares the subject's responses to two different sets of questions, one set of which is relevant to the investigation the other of which is not	Similar to the RIT but replaces the neutral irrelevant set with a comparison question set about unethical acts that most subjects would lie about to preserve their ego or reputation	Single set of multiple-choice questions, each with five or more possible answers, only one of which reflects the "true" state of affairs
Assumptions	Deceptive subjects will react more strongly to the relevant questions than to the irrelevant questions, while nondeceptive subjects will show no difference.	Deceptive subjects will react more strongly to the relevant questions than to the comparison questions, while nondeceptive subjects will show no difference.	Subjects who are concealing information about the "true" item will react more strongly to that item compared to the accompanying fictitious ones.
Example	Comparison of responses to an irrelevant question ("Is your age equal to x?) to a relevant question ("Did you enter address y through the fire escape window?")	Comparison of responses to a relevant question ("Did you enter address y through the fire escape window?") with a comparison question ("Have you ever removed printer paper from your place of work and taken it home?")	The assailant was beaten with a (*i*) hammer, (*ii*) pool cue, (*iii*) baseball bat, (*iv*) rock, (*v*) tire iron.
Strengths	Applicable across a variety of investigations requiring a polygraph examination	Applicable across a variety of investigations requiring a polygraph examination	Strong scientific (both theoretical and empirical) support for the approach with a variety of sensors
Limitations	Weak scientific (both theoretical and empirical) support for the approach with a variety of sensors	Weak scientific (both theoretical and empirical) support for the approach with a variety of sensors	Applicable to only a narrow subset of investigations

CURRENT LIMITATIONS OF TECHNOLOGICAL APPROACHES

Despite its widespread use, the polygraph remains a controversial and hotly debated approach to credibility assessment. It is relatively poorly understood outside the circle of its practitioners and a small group of academic researchers. This is not surprising given the secrecy that pervades the polygraph exam: details regarding the specifics of the signal processing performed within given polygraph instruments, as well as the previous results of polygraph examinations given to those known to have been spies, are hidden.[17] In its examination of polygraph-based personnel security screening, the National Research Council noted:

> There is a mystique surrounding the polygraph that may account for much of its usefulness: that is, a culturally shared belief that the polygraph device is nearly infallible. . . . In popular culture and media, the polygraph device is often represented as a magic mind-reading machine. These facts reflect the widespread mystique or belief that the polygraph test is a highly valid technique for detecting deception—despite the continuing lack of consensus in the scientific community about the validity of polygraph testing.[17]

The existing laboratory research into polygraphy does not support the public's belief in the machine's validity. The majority of the empirical research into the polygraph's effectiveness and accuracy has focused on the CQT design; a typical assessment of CQT accuracy rates found 75% for guilty subjects and only 50% (chance) for innocent subjects.[18] In other words, one out of four guilty subjects would escape detection, while half of the innocent subjects would be incorrectly evaluated as guilty. The RIT test is worse still because innocent subjects tend to react to the relevant question, particularly when the question is accusatory,[19] with one study reporting a false-positive rate greater than 70% with innocent subjects.[18] It should be noted that these accuracy rates, as low as they are, assume that the subjects are not using countermeasures; if subjects do use countermeasures, accuracy can drop even lower. Research has shown that the CQT in particular is vulnerable to the use of countermeasures, especially by those who have been trained to use them.[19] Not only are there serious questions about reliability, but the dominant protocols in deception detection itself are lacking a solid scientific foundation.[20] In response to these studies, a leading psychophysiologist has stated:

> It is evident that the field (a) is devoid of meaningful theory, (b) has failed to accumulate knowledge, (c) relies on studies of poor quality, (d) ignores evidence that contradicts the likely effectiveness of the technique, (e) continues to make claims that are unsubstantiated, and (f) makes claims that are difficult to believe given what we know about human psychophysiology.[19]

Generalizing results from laboratory experiments to real-world applications is a ubiquitous problem in applied sciences and is particularly problematic in credibility assessment research. For example, levels of motivation and stress are factors known to be important during credibility assessment, but laboratory experiments cannot typically achieve realistic motivations to deceive (even with monetary rewards), or induce the stresses that come with the possibility of passing or failing an actual polygraph test. Laboratory experiments have simulated a very narrow set of possible use cases, most involving mock theft or recognition of simple autobiographical details. In fact, only a single study included in the National Research Council's comprehensive report on the polygraph used data from a real screening situation.[17] Without more operationally realistic evaluation methodologies, it is difficult to envision useful and reliable innovation in credibility assessment, or any path to the admissibility of credibility assessments as legal evidence.

A significant weakness in the polygraph's approach is that the instrument is recording autonomic nervous system activity (e.g., changes in heart rate) that can be caused by a variety of psychological or physiological phenomena, such as generalized anxiety or pain. To be accurate, the polygraph needs to be able to detect a pattern that is uniquely linked to deception (given the context of a test where some of the potential confounds are controlled), but autonomic signals used by today's polygraph respond to far too general a set of conditions for such to be the case. A better understanding of the causal links between deceptive cognitive processes in the brain and the resultant activity in the autonomic nervous system could improve this situation,[7] but to date, there has been minimal research in this area.

If there is a bright spot in the polygraph world to contrast with the murky states of the CQT and RIT, it is the CIT. As the most recent of the major polygraph designs and the least controversial, the polygraph performs reasonably well under the CIT paradigm, identifying 85–90% of guilty subjects and often exonerating 100% of the innocent subjects in laboratory conditions, though some field-based studies have found lower accuracy rates.[18] Unlike the CQT and the RIT, the CIT is generally considered to be based on a more sound scientific foundation.[21] As a case in point, Japanese police polygraphers do not consider any polygraph tests other than the CIT sufficiently reliable, and consequently they administer the CIT exclusively in those investigations to which the polygraph is suitable.[22] It is notable that in Japan CIT test results can be admitted as evidence in court cases.

It is important to note that the CIT does not attempt to detect deception in its many forms per se, but rather focuses on whether or not a subject *recognizes* critical information (such as the specific murder weapon that was used). If a suspect who has denied involvement in such a murder is shown images of multiple weapons (e.g., several similar revolvers) and involuntarily (via his/her psychophysiological responses) demonstrates that he/she recognizes the correct one as the murder weapon, it can be inferred that the suspect has "guilty" knowledge of the event that would be known only to the perpetrator (and the investigators). Of course, if others might also have that knowledge from news reports or incidental contact, then the discovery that a subject recognizes the weapon would not be evidence of guilt. The CIT's psychological "recognition" signature appears to be more clearly and easily distinguished from the "no recognition" response than the corresponding "deception" versus "no deception" difference in responses for the CQT or RIT tests, although there are still significant gaps in the underlying research base.

However, two significant limitations have restricted the CIT's operational use. First, the CIT appears to be vulnerable to countermeasures, perhaps even more so than the CQT. For example, a countermeasure user could change their reactivity (by, for example, stepping on a thumbtack hidden inside their shoe) such that skin conductance responses to nonrelevant items are larger than their responses to relevant items. Countermeasures that inhibit a user's responses (e.g., by pharmaceuticals that depress the general emotional responses of the user) or increase the responses to all stimuli (maximizing all of the responses) can also be effective.[23]

More importantly, the opportunities to use the CIT are limited to situations in which the examiner knows specific details of the incident under investigation that would only be known by the guilty party and not by other innocent individuals. Unfortunately, many real-world cases do not provide suitable opportunities, perhaps as few as 15%.[18] An energetic, competitive press, along with a 24-hour news cycle driving a demand for more detailed information, tends to defeat attempts by law enforcement to keep critical facts of cases private (especially in the United States). The Japanese authorities who rely exclusively on the CIT have worked to make the CIT relevant in more cases by assigning a polygrapher to the crime scene to evaluate particular details before they can become public knowledge.[22] However, it is not clear how the Japanese success in employing the CIT could be extended to other applications that are of major concern to the U.S. intelligence community, such as the personnel screenings that account for tens of thousands of polygraph examinations each year. Any improvement in the accuracy of these methods would be of value, but it is difficult to imagine how the CIT, which is driven by specific details, could be applied to the multiple-issue, nonspecific nature of a personnel screening examination covering years of an individual's life.

FUTURE OPPORTUNITIES

At the request of the Department of Justice, the National Research Council (the research arm of the National Academies of Sciences, Engineering, and Medicine) conducted a review of the scientific basis underlying polygraphy, and the results were published in 2003. The report recognized the compelling national security need for credibility assessment but was critical of the current state of polygraphy, stating, "Its accuracy in distinguishing actual or potential security violators from innocent test takers is insufficient to justify reliance on its use in employee security screening in federal agencies."[17] To meet the present and likely future security needs, the National Research Council called for a renewed emphasis on credibility assessment research:

> We recommend an expanded research effort directed at methods for detecting and deterring major security threats, including efforts to improve techniques for security screening. . . . The research program should follow accepted standards for scientific research, use rules and procedures designed to eliminate biases that might influence the findings, and operate under normal rules of scientific freedom and openness to the extent possible while protecting national security).[17]

Since the publication of the report, there has been a small but significant effort to apply recent developments in the neurosciences, particularly cognitive neuroscience and social neuroscience, to advancing the credibility assessment state of the art. This research has been focused along two parallel but complementary avenues: (*i*) basic research directed toward building a better understanding of the cognitive and neural processes underlying deception in the human brain and nervous systems, and (*ii*) applied research attempting to develop better tools for detecting and identifying psychophysiological signatures of deception. A brief discussion of some of the applicable basic research and a recommended approach for moving forward is already available;[7] the remainder of this article focuses on the applied tools research efforts.

It should be noted that commercial entities have attempted to field neuroscience-based tools for credibility assessment. For example, at least two firms (No Lie MRI, Inc., and Cephos Corporation) have marketed functional magnetic resonance imaging (fMRI)-based approaches, while Brain Fingerprinting Laboratories has promoted a proprietary electroencephalography (EEG)-based approach.[7] While new tools are eagerly anticipated by the national security community, the proprietary nature of commercial products can interfere with a full, detailed examination of their capabilities and weaknesses. Scientific and government reviews of these commercial systems have been generally critical of their applicability or maturity; for example, a Government Accounting Office report on "brain fingerprinting" found little support for its use from security agencies and noted recommendations from scientific advisors that further research was necessary before relying on the system in actual field use.[24]

Many of the neuroscience tools and techniques that are key to basic cognitive neuroscience research are less likely to be useful as field-employable tools for credibility assessment. For example, fMRI has an excellent capability for resolving the spatial locations of metabolic changes in the brain (a marker of neural activity levels) but requires a 10-ton superconducting magnet and an electromagnetically shielded room to sense and record its signals accurately. In addition, some individuals cannot be safely scanned in an fMRI magnet because of the intense magnetic field. The same problems of size and expense apply to magnetoencephalography, whose temporal resolution is far superior to that of fMRI. A near-infrared spectroscopy instrument is relatively portable and inexpensive but is less capable and much more limited than fMRI or magnetoencephalography.

Among the neural tools relevant for credibility assessment in the field, the most extensive work has studied the use of EEG (see Box 1). The time-varying EEG signals reflect the electrical potential changes that accompany information processing in the brain as measured by sensors on the scalp. EEG waveforms contain underlying frequency components that can be correlated with brain states, such as when a person is sleeping, and can be used to diagnose certain brain-related abnormalities. A more recent development has been the capability to distinguish and classify brief, nonperiodic activity that results from a particular stimulus. When a stimulus (a picture, written text, etc.) is presented to an individual, a brief electrical transient occurs (after a short and characteristic time delay) that can be separated from longer-term state-related waveforms by using appropriate signal-processing algorithms. This transient evoked by a particular stimulus or averaged over a series of similar stimuli is referred to as the event-related potential (ERP). A given ERP is conventionally labeled as positive-going (P) or negative-going (N) along with the approximate delay (in milliseconds) after stimulus onset at which the transient typically occurs (for example, P300 or N400). Modern digital signal-processing techniques, combined with the rapid recent growth in computational power available in small computers, have made portable and relatively inexpensive EEG/ERP instrumentation a practical reality.

The most commonly studied ERP is known as the P300. This ERP has been used as an indication of familiarity, and it is particularly relevant for recognition detection in the context of concealed information testing. Commonly referred to as the "oddball" response, the P300 is a large positive deflection in the EEG response that follows the presentation of a stimulus, typically between 300 and 900 ms. The size of the P300 can be influenced by a variety of factors, including stimulus salience, stimulus relevance to a task, cognitive load, probability of occurrence, and others.[25–31] But the P300 is most consistently elicited by an oddball stimulus: a rarely presented stimulus that qualitatively differs from other recent stimuli. The response is thought to reflect the neural activity generated when a mental expectation or prediction is violated or a significant change is registered by the brain. If the proportion of stimuli that are irrelevant, unfamiliar, or insignificant to a subject is kept high, the rare stimulus that is relevant, familiar, or salient to a subject can be expected to evoke a P300 response, as illustrated in Fig. 3.[32]

Other ERPs are also thought to index familiarity within some experimental contexts, beginning around 200 ms following stimulus onset and lasting until 800 ms and longer. These include the N250,[33,34] the N400,[35] and the P600.[36] Although these additional ERP components have been linked to familiarity and recognition, the vast majority of EEG-based CIT (EEG-CIT) studies have focused exclusively on the P300.

BOX 1. ELECTROENCEPHALOGRAPHY

WHERE DO EEG SIGNALS COME FROM?

Neurons are constantly communicating with one another through electrochemical changes. The largest of these changes involve neuronal spikes, which are particularly large (but brief) voltage discharges that are propagated from neuron to neuron as the brain processes information. The voltage discharges of individual neurons are extremely small, but when a sufficient number of neurons spike synchronously, larger voltage changes are produced and travel through the fluids within and surrounding the brain, and then through the skull and skin, and can ultimately be detected by electrodes on the scalp (Fig. 2).

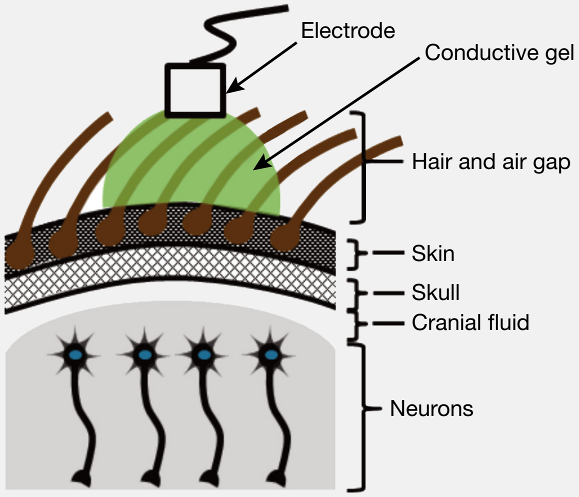

Figure 2. Sources and measurement of EEG signals.

HOW CAN WE MEASURE EEG SIGNALS?

Despite reflecting the activity of millions of neurons, these "brainwave" voltages are still quite small and challenging to detect. To improve the contact between electrodes and the scalp, EEG sensors typically use a conductive gel that bridges the electrode–skin junction (although there has been significant progress in recent years in the development of "dry" EEG electrodes that do not use such gels). Even with gels, EEG recordings are still subject to noise and have a relatively low signal-to-noise ratio. Two common methods are used to overcome these low signal-to-noise ratios: (*i*) analyzing EEG power in specific frequency bands (e.g., activity in the 8- to 12-Hz range, or alpha, has been associated with changes in attention or mental engagement), and (*ii*) event-related averaging, in which a stimulus or event is repeated multiple times so that the EEG signal can be averaged to produce an ERP.

WHAT ARE ERPs?

Several different characteristic ERPs have been identified. One class is the ERN, which is a negative voltage potential often observed when the brain detects a recent error or mistake. The P300 ERP (so named because it manifests as a positive voltage deflection roughly 300 ms following a stimulus) is sometimes called the oddball response because it tends to be strongest when the brain has been exposed to a stimulus that stands out or differs relative to other recent stimuli (Fig. 3). For example, a P300 could be evoked by a high-pitched tone that follows a repetitive sequence of low tones or, similarly, by a picture that stands out in an unexpected way compared to other recent pictures.

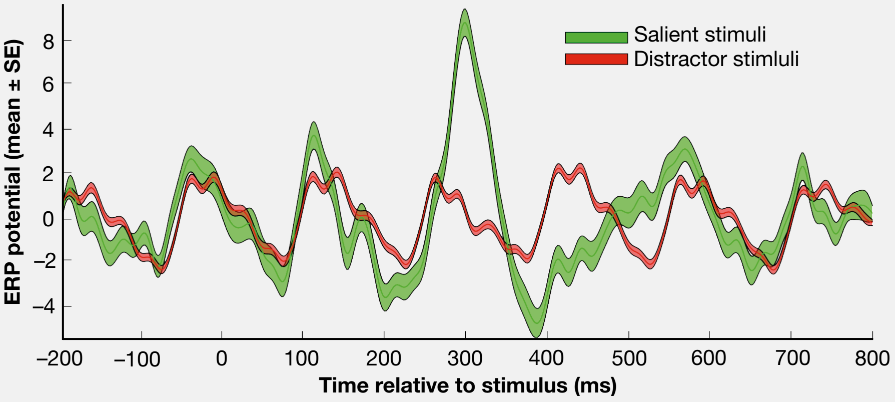

Figure 3. Example of an ERP signal (P300).

Several different protocols have been developed for eliciting ERPs for use in concealed information testing: the Oddball Protocol (OP), which is the most straightforward; the Three Stimulus Protocol (3SP), which attempts to prevent inattention-based countermeasures; and the Complex Trial Protocol (CTP), which attempts to improve sensitivity and further reduce susceptibility to countermeasures.

1. In the OP there are two sets of stimuli: (*i*) probes, which are relevant, familiar, or salient stimuli; and (*ii*) irrelevant, unfamiliar, or insignificant stimuli. Probe stimuli are presented rarely (oddballs) as compared to irrelevant stimuli. The test assumes that guilty subjects will generate a clear, significant P300 ERP in response to familiar stimuli. The simplicity of the OP unfortunately increased its susceptibility to countermeasures, such as an inattentive subject.

2. The 3SP modifies the OP to prevent inattention-based countermeasures by adding a third class of stimuli to verify the attentiveness of the subject: target items that are known to be familiar to the subject (e.g., his wife's name). Here the subject is tasked with making a behavioral response (e.g., a button press) when each target stimulus appears to verify that the subject is paying close attention to the stimuli stream. The test assumes that guilty subjects will generate a clear, significant P300 ERP in response to both relevant and target stimuli.[37,38]

3. The CTP modifies the 3SP to address both sensitivity to concealed information and vulnerability to countermeasures.[39,40] As shown in Fig. 4 below, each CTP trial is divided into two parts: in the first part, a probe or irrelevant stimulus is presented (similar to the 3SP) followed by a mask period during which the subject is instructed to simply press a button following every stimuli (e.g., an "I saw it" response); in the second part, a target or nontarget stimulus is presented, also followed by a mask period. Prior to testing, the subject is trained to identify target stimuli (e.g., text of a specific color) and is instructed to make a forced choice during the second part of the complex trial: press one button for a target and a different button for a nontarget. In this way, the ERP evoked by the first stimulus is used to evaluate probe recognition, the response time to the first stimulus is used to monitor for countermeasures, the ERP evoked by the target stimuli can be used as a reference for ERPs to highly salient stimuli, and the response accuracy to the second stimulus is used to verify sufficient engagement in the task. As an additional test for attentional engagement, participants are also briefed prior to the evaluation that periodically the process will be paused and they will be quizzed regarding the specific identity of the first stimuli (with performance penalties for frequent mistakes). Under laboratory conditions, several evaluations of the CTP have suggested that it is resistant (although not immune) to countermeasures while also being highly accurate in detecting recognition of known stimuli.

A comprehensive meta-analysis that synthesized results across 229 studies[41] recently summarized decades of CIT research using these protocols with both EEG-based and traditional autonomic measures. The principal finding was that CIT using EEG/ERP performed best. Both EEG and skin-conductance measures were superior to heart rate measures or respiration-related measures.

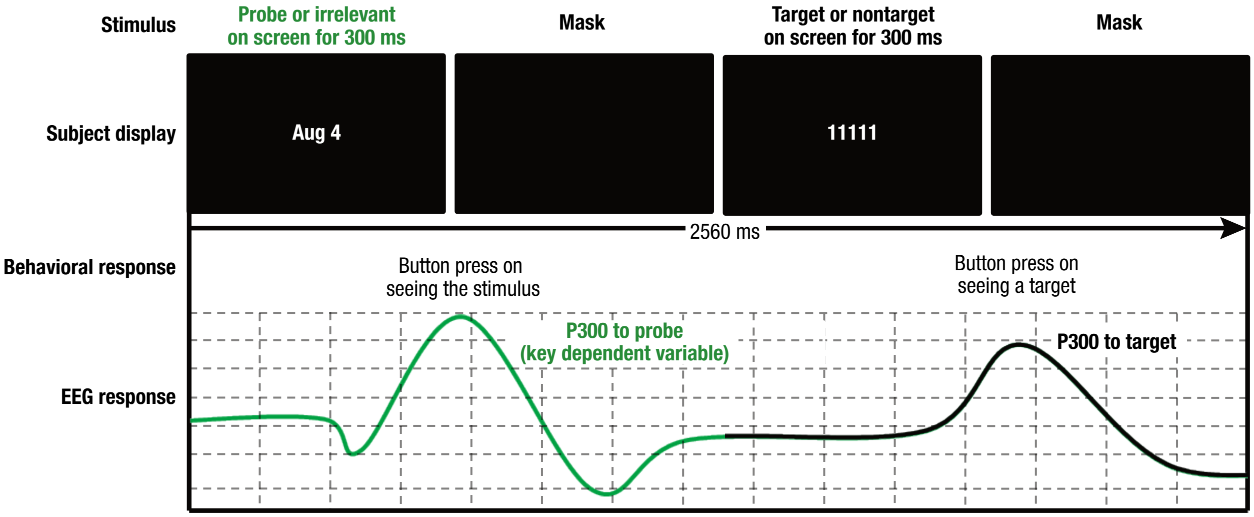
Stimulus
Probe or irrelevant on screen for 300 ms
Mask
Target or nontarget on screen for 300 ms
Mask
Subject display
Aug 4
11111
2560 ms
Behavioral response
Button press on seeing the stimulus
Button press on seeing a target
EEG response
P300 to probe (key dependent variable)
P300 to target

Figure 4. CTP, adapted from Refs. 39 and 40. Each trial consists of two stimuli: The first stimulus is either a probe or an irrelevant item, always followed by a button press by the subject. The button press can be used to look for countermeasure usage, while the ERP following the stimuli is used to detect concealed information or salience of the probe. The second stimulus requires a target/nontarget decision and a button press and is used to ensure engagement in the task.

This meta-analysis clearly demonstrates the potential of EEG-CIT, but important questions remain:

- How susceptible is EEG-CIT to countermeasures? A recent review[40] focused explicitly on the CTP found the technique to be quite resistant, although not immune; sensitivity and specificity scored above 90%.
- Does the amount of time that passes between a deceptive act and a credibility assessment test affect the test's accuracy? One investigation[42] found EEG-CIT effective even when performed 1 month after a mock crime scenario.
- Will outsider knowledge about the event under investigation impair the test? One study found that a 69% false-alarm rate was observed when innocent participants were informed about probe items prior to the testing, as compared to a 14% false-alarm rate for naive innocent participants.[43] Still, basic neuroscience research has demonstrated that responses associated with personally familiar or recalled content can be distinguished from responses associated with incidental or merely recognized content.[36]
- How do traditional and EEG-based approaches fare outside the laboratory in more ecologically valid testing scenarios?

CONCLUSIONS

The DoD and intelligence community have critical operational needs for accurate and reliable means of detecting deception. Existing methods, while in many cases better than chance, are nevertheless insufficient to fulfill those needs. The polygraph has been in use in various forms since its development early in the 20th century, but its performance (even in controlled laboratory settings) has not matched the faith practitioners have in the approach. After conducting a 19-month study of the polygraph at the request of the U.S. Department of Energy (DOE), the National Research Council issued the following conclusion:

> Polygraph testing yields an unacceptable choice for DOE employee security screening between too many loyal employees falsely judged deceptive and too many major security threats left undetected. Its accuracy in distinguishing actual or potential security violators from innocent test takers is insufficient to justify reliance on its use in employee security screening in federal agencies.[17]

Nevertheless, federal agencies will have to continue to rely on polygraph testing for employee security screening and/or asset validation until better alternatives become available.

From this review, we conclude that there is substantial room for improvement in all of the current methods for credibility assessment. The polygraph-based CIT demonstrates the best performance but is limited in its practical application (especially in a security screening role) and appears to be the most vulnerable to countermeasures. Research aimed at reducing or eliminating these limitations would be especially valuable. In the near term, EEG could change the focus from measuring general autonomic signals to recording specific brain-based measures whose role in deception is better understood. Beyond the near-term applications of EEG protocols and processing, a substantial program of research to better understand the nature of deceptive cognition and recognition processes, the psychophysiological and neural bases of these constructs, and how advanced sensors, signal processing, and computation would be necessary to create truly effective means of testing and validating personnel credibility. Finally, better evaluation methodologies will be necessary to determine which techniques and technologies should be considered truly effective.

REFERENCES

[1] Cooper, J. R., *Curing Analytic Pathologies: Pathways to Improved Intelligence Analysis*, Center for the Study of Intelligence, Central Intelligence Agency, Washington DC (2005).

[2] Ford, C. V., *Lies! Lies! Lies! The Psychology of Deceit*, American Psychiatric Press, Washington DC (1996).

[3] DePaulo, B. M., and Pfeifer, R. L., "On-the-Job Experience and Skill at Detecting Deception," *J. Appl. Soc. Psychol.* **16**(3), 249–267 (1986).

[4] Ekman, P., and O'Sullivan, M., "Who Can Catch A Liar?" *Am. Psychol.* **46**(9), 913–920 (1991).

[5] Ekman, P., O'Sullivan, M., and Frank, M. G., "A Few Can Catch a Liar," *Psychol. Sci.* **10**(3), 263–266 (1999).

[6] Department of Defense Directive 5210.48, *Credibility Assessment (CA) Program*, U.S. Department of Defense, Washington, DC (24 Apr 2015).

[7] Happel, M. D., "Neuroscience and the Detection of Deception," *Rev. Policy Res.* **22**(5), 667–685 (2005).

[8] Heckman, K., and Happel, M., "Mechanical Detection of Deception: A Short Review," *Educing Information: Interrogation: Science and Art*, R. Swenson (ed.), National Defense Intelligence College, Washington, DC, pp. 63–94 (2006).

[9] Fisher, R. P., and Geiselman, R. E., *Memory-Enhancing Techniques for Investigative Interviewing: The Cognitive Interview*, Charles C. Thomas Publisher, Springfield, IL (1992).

[10] Memon, A., Meissner, C. A., and Fraser, J., "The Cognitive Interview: A Meta-Analytic Review and Study Space Analysis of the Past 25 Years," *Psychol. Public Policy Law* **16**(4), 340–372 (2010).

[11] Bensi, L., Nori, R., Gambetti, E., and Giusberti, F., "The Enhanced Cognitive Interview: A Study on the Efficacy of Shortened Variants and Single Techniques," *J. Cognit. Psychol.* **23**(3), 311–321 (2011).

[12] Dando, C., Wilcock, R., Milne, R., and Henry, L., "A Modified Cognitive Interview Procedure for Frontline Police Investigators," *Appl. Cognit. Psychol.* **23**(5), 698–716 (2009).

[13] Paulo, R. M., Albuquerque, P. B., and Bull, R., "The Enhanced Cognitive Interview: Towards a Better Use and Understanding of This Procedure," *Int. J. Police Sci. Manage.* **15**(3), 190–199 (2013).

[14] Grubin, D., "The Polygraph and Forensic Psychiatry," *J. Am. Acad. Psychiatry Law Online* **38**(4), 446-451 (2010).

[15] McGrath, R. J., Cummings, G. F., Hoke, S. E., and Bonn-Miller, M. O., "Outcomes in a Community Sex Offender Treatment Program: A Comparison Between Polygraphed and Matched Non-Polygraphed Offenders," *Sex. Abuse* **19**(4), 381–393 (2007).

[16] Honts, C., "The Psychophysiological Detection of Deception," *The Detection of Deception in Forensic Contexts*, P. A. Granhag and L. A. Strömwall (eds.), Cambridge University Press, Cambridge, UK, pp. 103–123 (2004).

[17] National Research Council, *The Polygraph and Lie Detection*, National Academies Press, Washington, DC (2003).

[18] Iacono, W., "The Detection of Deception," *Handbook of Psychophysiology* (2d Ed.), J. Cacioppo, L. Tassinary, and G. Berntson (eds.), Cambridge University Press, Cambridge, UK, pp. 772–793 (2000).

[19] Iacono, W. G., "Effective Policing: Understanding How Polygraph Tests Work and Are Used," *Crim. Justice Behav.* **35**(10), 1295–1308 (2008).

[20] Ben-Shakhar, G. A., "Critical Review of the Control Question Test (CQT)," *Handbook of Polygraph Testing*, M. Kleiner (ed.), Academic Press, San Diego, pp. 103–126 (2002).

[21] Ben-Shakhar, G., and Elaad, E., "The Guilty Knowledge Test (GKT) as an Application of Psychophysiology: Future Prospects and Obstacles," *Handbook of Polygraph Testing*, M. Kleiner (ed.), Academic Press, San Diego, pp. 87–102 (2002).

[22] Nakayama, M., "Practical Use of the Concealed Information Test for Criminal Investigation in Japan," *Handbook of Polygraph Testing*, M. Kleiner (ed.), Academic Press, San Diego, pp. 49–86 (2002).

[23] Honts, C. R., and Amato, S. L., "Countermeasures," in *Handbook of Polygraph Testing*, M. Kleiner (ed.), Academic Press, San Diego, pp. 251–264 (2002).

[24] GAO, *Investigative Techniques: Federal Agency Views on the Potential Application of "Brain Fingerprinting,"* US General Accounting Office, Washington, DC (2001).

[25] Berlad, I., and Pratt, H., "P300 in Response to the Subject's Own Name," *Electroencephalogr. Clin. Neurophysiol.* **96**(5), 472–474 (1995).

[26] Castro, A., and Díaz, F., "Effect of the Relevance and Position of the Target Stimuli on P300 and Reaction Time," *Int. J. Psychophysiol.* **41**(1), 43–52 (2001).

[27] Donchin, E., and Coles, M., "Is the P300 Component a Manifestation of Context Updating?" *Behav. Brain Sci.* **11**(3), 357–374 (1988).

[28] Ford, J., Roth, W., and Kopell, B., "Auditory Evoked Potentials to Unpredictable Shifts in Pitch," *Psychophysiol.* **13**(1), 32–39 (1976).

[29] Polich, J., "P300, Probability and Interstimulus Interval," *Psychophysiol.* **27**(4), 396–403 (1990).

[30] Polich, J., "Updating P300: An Integrative Theory of P3a and P3b," *Clin. Neurophysiol.* **118**(10), 2128–2148 (2007).

[31] Rosenfeld, J. P., Biroschak, J. R., Kleschen, M. J., and Smith, K. M., "Subjective and Objective Probability Effects on P300 Amplitude Revisited," *Psychophysiol.* **42**(3), 356–359 (2005).

[32] Abootalebi, V., Moradi, M., and Khalilzadeh, M., "A New Approach for EEG Feature Extraction in P300-Based Lie Detection," *Comput. Methods Programs Biomed.* **94**(1), 48–57 (2009).

[33] Miyakoshi, M., Nomura, M., and Ohira, H., "An ERP Study on Self-Relevant Object Recognition," *Brain Cogn.* **63**(2), 182–189 (2007).

[34] Sun, D., Chan, C. C. H., and Lee, T. M. C., "Identification and Classification of Facial Familiarity in Directed Lying: An ERP Study," *PloS One* **7**(2), e31250 (2012).

[35] Sun, D., Lee, T. M. C., and Chan, C. C. H., "Unfolding the Spatial and Temporal Neural Processing of Lying About Face Familiarity," *Cereb. Cortex* **25**(4), 927–936 (2015).

[36] Touryan, J., Gibson, L., Horne, J. H., and Weber, P., "Real-Time Measurement of Face Recognition in Rapid Serial Visual Presentation," *Front. Psychol.* **2**(42), 1–8 (2011).

[37] Bowman, H., Filetti, M., Janssen, D., Su, L., Alsufyani, A., and Wyble, B., "Subliminal Salience Search Illustrated: EEG Identity and Deception Detection on the Fringe of Awareness," *PloS One* **8**(1), e54258 (2013).

[38] Bowman, H., Filetti, M., Alsufyani, A., Janssen, D., and Su, L., "Countering Countermeasures: Detecting Identity Lies by Detecting Conscious Breakthrough," *PloS One* **9**(3), e90595 (2014).

[39] Rosenfeld, J. P., Labkovsky, E., Winograd, M., Lui, M. A., Vandenboom, C., and Chedid, E., "The Complex Trial Protocol (CTP): A New, Countermeasure-Resistant, Accurate, P300-Based Method for Detection of Concealed Information," *Psychophysiol.* **45**(6), 906–919 (2008).

[40] Rosenfeld, J. P., Hu, X., Labkovsky, E., Meixner, J., and Winograd, M. R., "Review of Recent Studies and Issues Regarding the P300-Based Complex Trial Protocol for Detection of Concealed Information," *Int. J. Psychophysiol.* **90**(2), 118–134 (2013).

[41] Meijer, E. H., Smulders, F. T. Y., Merckelbach, H. L. G. J., and Wolf, A., "The P300 Is Sensitive to Concealed Face Recognition," *Int. J. Psychophysiol.* **66**(3), 231–237 (2007).

[42] Hu, X., Hegeman, D., Landry, E., and Rosenfeld, J., "Increasing the Number of Irrelevant Stimuli Increases Ability to Detect Countermeasures to the P300-Based Complex Trial Protocol for Concealed Information Detection," *Psychophysiol.* **49**(1), 85–95 (2012).

[43] Winograd, M. R., and Rosenfeld, J. P., "The Impact of Prior Knowledge from Participant Instructions in a Mock Crime P300 Concealed Information Test," *Int. J. Psychophysiol.* **94**(3), 473-481 (2014).

Mark D. Happel, Air and Missile Defense Sector, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Mark D. Happel is an information scientist in APL's Air air Missile Defense Sector. He holds a B.S.E.E. in electrical engineering from the United States Naval Academy, an M.S.E. in electrical engineering from the University of Central Florida, and a D.Sc. in computer science from the George Washington University. He develops signal processing and simulation software for missile defense applications and performs research in machine learning and cognitive neuroscience. Prior to working for APL, he led engineering, neuroscience, and public policy research and development efforts at federally funded research and development centers and commercial firms; while in the U.S. Navy, he served as a nuclear-trained submarine officer and attained the rank of Commander, U.S. Navy Reserve. He has taught graduate-level courses in artificial intelligence, machine learning, cognitive modeling, and software engineering as an adjunct faculty member for both the Johns Hopkins University and the George Washington University. He is a member of the Society for Neuroscience and serves on the executive committee of the annual IEEE-sponsored Applied Imagery Pattern Recognition workshop. His e-mail address is mark.happel@jhuapl.edu.

Jason A. Spitaletta, Asymmetric Operations Sector, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Jason Spitaletta is a psychologist in APL's Asymmetric Operations Sector and a major in the U.S. Marine Corps Reserve. His primary research experience is in applied, experimental, and political psychology and cognitive neuroscience; he also has operational experience in psychological operations/military information support operations and intelligence assignments in the U.S. Marine Corps as well as the joint and special operations communities. He is an adjunct faculty member at the National Intelligence University. He holds a bachelors' degree in biochemistry from Franklin & Marshall College, a master's degree in human factors from Embry-Riddle Aeronautical University, and a master's degree and a Ph.D. in applied experimental psychology from Catholic University. He also holds a graduate certificate from Stanford University's Summer Institute for Political Psychology. His e-mail address is jason.spitaletta@jhuapl.edu.

Eric A. Pohlmeyer, Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Eric A. Pohlmeyer is a member of the Senior Professional Staff in the Intelligent Systems Group in APL's Research and Exploratory Development Department. He received a B.S. in mechanical engineering from the University of Cincinnati and M.S. and Ph.D. degrees in biomedical engineering from Northwestern University. When he is not writing about himself in the third person, Dr. Pohlmeyer does research in neural signal processing and decoding, motor function and brain–machine/brain–computer interfaces (BMIs/BCIs). He has developed BMIs for controlling robotic arms using reinforcement learning algorithms, BMIs that can decode desired hand movements from the brain and translate them into electrical stimulation of paralyzed muscles to restore wrist function, and BMIs for controlling flight stimulators. He has also worked with EEG-based neural interfaces, in particular with methods for detecting interest in the brain following visual stimuli (which can be used to help individuals sort through large-scale image databases), and methods for measuring cognitive workload in pilots. His e-mail address is eric.pohlmeyer@jhuapl.edu.

Grace M. Hwang, Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Grace M. Hwang is the Program Manager for the Neurological Health and Human Performance Program in APL's Research and Exploratory Development Department. She holds an M.S. in civil and environmental engineering from the Massachusetts Institute of Technology and an M.S. and a Ph.D. in biophysics and structural biology. Her dissertation research focused on the analysis of human electroencephalography (EEG) to understand the neural basis of visual and verbal memory; she developed novel nonparametric statistical techniques for high-dimensional data for across- and within-subject multifactorial analysis. She leads a talented team of scientists, engineers, and clinicians in research efforts focused on understanding the structure and function of the brain in order to develop technologies to improve cognitive performance, make strides against neurobiological disease and injury, and develop novel brain–computer interfaces to transform how humans interact with each other and the world around us. She has extensive experience in experimental design, computational cognitive neuroscience, biophysics, biosensors, biomarker discovery, and optical spectroscopy. She contributed to the state of the practice research, writing, and review of this article. Her e-mail address is grace.hwang@jhuapl.edu.

Ariel M. Greenberg, Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Ariel M. Greenberg is a senior research scientist in APL's Intelligent Systems Center; Assistant Editor-in-Chief of the *Johns Hopkins APL Technical Digest* (http://www.jhuapl.edu/techdigest/); Steering Committee Co-chair of the International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation (http://sbp-brims.org/); and an instructor in APL's Strategic Education Program. In the spirit of cells-to-societies transdisciplinary research, his work addresses topics in psychophysiology and computational social science, toward applications in defense, intelligence, and health. Ariel received degrees in biology and electrical engineering from University of Maryland, College Park. His e-mail address is ariel.greenberg@jhuapl.edu.

Clara A. Scholl, Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Clara A. Scholl is a neural signal processing engineer in APL's Intelligent Systems Center. Clara completed a B.A. in physics at Kalamazoo College and a Ph.D. in neuroscience at Georgetown University, where she studied the temporal dynamics of visual object recognition in cortex. At APL, she leads research efforts in cognitive load monitoring, brain–computer interfaces, and neural decoding, as well as technology development for neural signal extraction and neural stimulation. She contributed background on applications of EEG and cognitive load monitoring for this article. Her e-mail address is clara.scholl@jhuapl.edu.

Michael Wolmetz, Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Michael Wolmetz is a project manager in APL's Research and Exploratory Development Department. He received a B.S. in computer science from Yale University and a Ph.D. in cognitive science from Johns Hopkins University. He leads several projects focusing on defense, intelligence, and clinical applications for neuroimaging and cognitive assessment. His e-mail address is michael.wolmetz@jhuapl.edu.