Cognitive and Affective Probing: A Tutorial and Review of Active Learning for Neuroadaptive Technology

Laurens R. Krol[1,*], Pim Haselager[2], and Thorsten O. Zander[3]

[1]Biological Psychology and Neuroergonomics, Technische Universität Berlin, Berlin, Germany
[2]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
[3]Zander Laboratories B.V., Amsterdam, The Netherlands
[*]Correspondence: `lrkrol@gmail.com`

Abstract

Objective: The interpretation of neurophysiological measurements has a decades-long history, culminating in current real-time brain-computer interfacing (BCI) applications for both patient and healthy populations. Over the course of this history, one focus has been on the investigation of cortical responses to specific stimuli. Such responses can be informative with respect to the human user's mental state at the time of presentation. An ability to decode neurophysiological responses to stimuli in real time becomes particularly powerful when combined with a simultaneous ability to autonomously produce such stimuli. This allows a computer to gather stimulus-response samples and iteratively produce new stimuli based on the information gathered from previous samples, thus acquiring more, and more specific, information. This information can even be obtained without the explicit, voluntary involvement of the user.

Approach: We define cognitive and affective probing, referring to an application of active learning where repeated sampling is done by eliciting implicit brain responses. In this tutorial, we provide a definition of this method that unifies different past and current implementations based on common aspects. We then discuss a number of aspects that differentiate various possible implementations of cognitive probing.

Main results: We argue that a key element is the user model, which serves as both information storage and basis for subsequent probes. Cognitive probing can be used to continuously and autonomously update this model, refining the probes, and obtaining increasingly detailed or accurate information from the resulting brain activity. In contrast to a number of potential advantages of the method, cognitive probing may also pose a threat to informed consent, our privacy of thought, and our ability to assign responsibility to actions mediated by the system.

Significance: This tutorial provides guidelines to both implement, and critically discuss potential ethical implications of, novel cognitive probing applications and research endeavours.

1 Introduction

One essential skill required of disc jockeys (DJs) is the ability to "read the crowd". In order to maximise the audience's enjoyment, a DJ may play different songs and gauge

the reaction they evoke (Gates, Subramanian, & Gutwin, 2006). Depending on this reaction, the DJ decides whether or not to end the current song early, and what to play next. This is an iterative process that continues to optimise with each played song, the goal being maximum audience participation, or indeed, maximum audience happiness.

The feedback from the audience is essential input in this process. However, the audience members themselves are not necessarily aware of providing any—their responses are mostly automatic, even reflexive. Hardly would one deliberate whether or not the current chords are sufficiently engaging; rather, the rhythm and mood of the music will simply induce the audience members to express themselves in a certain way, and it is this expression that the DJ will read and use as input.

In this tutorial, we will discuss a situation analogous to the one above, in the context of human-computer interaction (HCI): An adaptive process where a computer presents something that automatically evokes a response from the user, which the computer can then interpret and use for further processing. We will discuss the methods involved, how they relate to past and current applications, and what aspects must be considered going forward with this technology.

Let us first discuss what an "automatic" response may be in the context of HCI. In traditional HCI, all input that we give to the computer is based on goal-oriented, explicit, voluntary actions. By *explicit communication*, we mean any intentional action performed for the purpose of communicating specific content. For example, to communicate information to a computer, we must actively type, say, click, move, swipe, or select things in a corresponding fashion, or nothing will happen.

There are technologies that give a computer access to more information about the user than what is explicitly communicated. In virtual reality, moving our arms and legs provides explicit positional information to the computer, but, going beyond that, full-body motion tracking can also be used to analyse these movements, or the resulting postures, for additional information that was not meant to be communicated (Kleinsmith, Bianchi-Berthouze, & Steed, 2011). Your movements may reveal that you are happy or sad, or perhaps drunk. Similarly, for example, facial recognition can be used to assess emotions (Janssen et al., 2013), and passive brain-computer interfaces can analyse our brain activity in order to supply implicit input to a computer (Zander & Kothe, 2011; Krol, Andreessen, & Zander, 2018). By *implicit communication*, we mean that information is acquired from output that was not intended to communicate that information. Here, we focus on methods that target our brains—the primary seat of cognition and conscious experience—as they are uniquely positioned to infer information about us in such a way.

Real-time detection of mental states has been demonstrated as early as the 1970s (Vidal, 1977), and since then, further research has shown that BCI methodology can be used to detect a range of different patterns of brain activity, including some that reflect specific cognitive and affective processes (e.g., Donchin, 1981; Hillyard & Anllo-Vento, 1998; Yeung & Sanfey, 2004). These mental processes occur naturally during our everyday lives. So-called *passive* BCI systems (Zander & Kothe, 2011; Krol et al., 2018) target such naturally-occurring mental states in order to provide *implicit input* to a computer (Schmidt, 2000; Zander, Brönstrup, Lorenz, & Krol, 2014). In other words, passive BCI systems rely on brain activity that was not *intended* by the user to provide information to those systems, but is nonetheless used as such.

It should be noted, however, that our current ability to discern different mental states, and thus our ability to infer specific information from brain activity, is limited. While research into "reading" mental states in novel ways continues to move forward

(e.g., Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016; Haynes, 2011; Senden, Emmerling, van Hoof, Frost, & Goebel, 2019), it is not currently possible to infer the propositional content of thought (Haselager & Mecacci, 2018). Thus, in many cases, additional information about the current situation is required in order to draw meaningful inferences from measured brain activity (Makeig, Gramann, Jung, Sejnowski, & Poizner, 2009; Zander & Jatzev, 2012). For example, when high workload is detected in a multi-tasking environment, additional information is required to make an optimal decision on which task to automate (Mühl, Jeunet, & Lotte, 2014; Krol, Freytag, Fleck, Gramann, & Zander, 2016). Similarly, when error processing is detected, it is important to know where that error originated (Falkenstein, Hoormann, Christ, & Hohnsbein, 2000; Mousavi & de Sa, 2019; Wirth, Dockree, Harty, Lacey, & Arvaneh, 2019). Determining the best way for a computer to sense, filter, and represent relevant situational aspects remains an open field of research with many different approaches (Perera, Zaslavsky, Christen, & Georgakopoulos, 2014).

In our earlier analogy, if an observer to the DJ's performance sees the crowd's reaction but cannot hear and does not know what music is currently playing, they will be none the wiser, until they obtain this extra piece of information. The DJ, on the other hand, doesn't only know what music is playing—she is the one who selected the music in the first place!

Similar to a DJ, in HCI contexts where a computer has control over relevant situational aspects as well as access to mental state measurements, it can also *actively induce an even in order to assess the user's brain response to it*. This allows it to *probe* the human user. Having a computer induce an event itself circumvents the need for elaborate post-hoc context sensing. Rather than waiting for events to occur naturally in order to learn the relation between events and mental states, the computer can purposefully generate specific events and gauge the user's implicit response. Doing this in an iterative fashion allows the computer to adaptively pursue a particular goal. As we will see, these events can also be "hidden", embedded in the ongoing interaction itself. Furthermore, considering that humans are generally unable to stop their brains from automatically processing perceived stimuli, the computer is effectively posing a question directly to the user's brain, potentially bypassing the user's explicit faculties.

In essence, this *cognitive probing* can be seen as a form of *active learning* (Settles, 2009). Active learning is a concept in machine learning where a learner, rather than simply accepting samples as they come, has control over which samples it learns from. Cognitive probing applies this concept to a computer learning from implicit responses to technological state changes, with the added notion that the learner in this case is not merely responsible for its own learning, but also for the effect its sampling has on the human. Inducing mental states is not a neutral act.

Learning itself can be the sole goal of the computer. However, additional goals may be pursued as well: much like the DJ's goal of maximum audience happiness, the computer, too, may attempt to find stimuli that maximise certain psychological states. As such, cognitive probing can be a powerful tool to realise closed-loop neuroadaptive technology. Neuroadaptive technology is technology that adapts itself based on implicit input of neurophysiological origin (Zander & Krol et al., 2016). The amount of input obtained by such technology, and the efficiency with which it is obtained, can be greatly increased using cognitive probing. In particular, the technology can iteratively adapt itself to the user based on implicit information gathered from probe-induced brain activity.

The use of passive BCIs for cognitive probing thus provides a potentially powerful

method for a computer to unobtrusively obtain information about its users. At the same time, it also represents a potentially dangerous and psychologically invasive method concerning mental privacy, and hence its application must be handled with due care and transparency (Mecacci & Haselager, 2019b).

In this tutorial, we first introduce and explain a general, broad definition of this method. We provide examples from the literature, pointing out that cognitive probing, as defined here, has already been used for at least three decades in different ways and different fields. With recent advances in neurophysiological sensing (Mullen et al., 2015; Zander et al., 2017) and neuroadaptive technology (Zander & Krol et al., 2016; Lorenz, Hampshire, & Leech, 2017), however, we believe it is prudent to now unify these methods previously considered disparate in a single framework, and to present a careful consideration of the different aspects involved. Therefore, a further section discusses a list of aspects, characteristics, and other considerations that are relevant to the use of cognitive probing in different applications, and to society at large. We conclude with a speculative vision of the future, highlighting possible research endeavours.

2   Cognitive Probing

At the basis of our proposed definition lies a single probe, where:

A cognitive (or affective) probe is a technological state change that is initiated or manipulated by the same technology that uses it to learn from the user's implicit, situational, cognitive or affective brain response to it.

Cognitive probing or *affective probing*, then, is the repeated use of probes in order to generate a model in which the learned information is registered.[1]

We use cognitive probing as the general term including both cognitive and affective probing; the term affective probing can be used in cases focusing exclusively on affect (Zander & Jatzev, 2009). Also note that the base definition refers to a single probe, i.e. a single state change eliciting a single response. Technically, meaningful learning can be done using a single probe: given sufficiently accurate tools, a single probe may be sufficient to provide adequate information. However, the main potential, and main focus, of the method lies in the adaptive, sequential use of multiple probes.

We will now discuss the definition's constituent elements in more detail.

Technological state change.   A technological state change is any action undertaken by a piece of technology. We use the term *technology* to collectively refer to any and all (connected) technological elements in any potential situation. With a technology's *state* being its specific current configuration, a *technological state change* is thus any change in configuration that the technology undergoes.

In neurophysiological experiments, the presentation of cues, stimuli, or feedback would be an example of a state change. In the example of an automated, neuroadaptive DJ, it could be a song that is started.

Initiated or manipulated.   Most technological state changes that a user will notice are user-generated, i.e. commanded by the user themselves, with the technology reacting

---

[1]The definitions given here for cognitive probe and cognitive probing supersede the earlier working definitions presented in Krol and Zander (2018).

to explicit input. Non-user-generated events include e.g. notifications of incoming e-mails or low battery. A cognitive probe refers to a state change whose occurrence, form, or timing was at least partially produced by the technology itself, and not by the user. This thus requires that the technology has some degree of control over its own state changes.

Technology can initiate a state change for the sole or primary purpose of obtaining a response, i.e. to produce a probe for the probe's sake. When a potentially informative state change occurs for other purposes, user-generated or otherwise, the system can manipulate it so as to serve as a probe, for example by adjusting its form or timing.

In the example started above, it is important that the neuroadaptive technology itself somehow influences the song, as opposed to a human user simply pressing "play". The manipulation can take different forms: among other options, it could be that the user requests music to be played in general, upon which the technology selects the song; it could be that the music has been pre-selected but the technology decides when to play it; or it could be a process chain initiated entirely by the technology itself. In any case, the technology does initiate or manipulate the technological state in question, as opposed to this being decided entirely by the human.

The user's implicit, situational, cognitive or affective brain response. Cognitive probing focuses on *brain responses*: brain activity (or a change in brain activity) that is elicited as a function of specific events. To be a cognitive probe, a technological state change must thus in one way or another elicit brain activity, or a change in ongoing brain activity, which can in turn be detected by the technology.

"Cognitive or affective" is spelled out here to distinguish it from more elementary brain activity such as the auditory brainstem response or steady-state visually evoked potentials. Since these responses do not inherently reflect cognitive or affective processing, they cannot be used for cognitive probing.

Furthermore, we limit this definition to information taken from the brain response that was not intended to be communicated to the technology—cognitive probing targets *implicit* human-computer communication.

"Situational" means the implicit brain response is automatically registered alongside parameters of the current situation. An ability to interpret ongoing brain activity taken in and of itself results merely in a stream of mental state assessments. In order to learn something from these mental states, they must be combined with additional information about the situation in which these mental states occurred. Cognitive probes ensure that at least one piece of situational information is available: the probe itself.

Further situational aspects can of course be included, dependent on the accuracy with which they can be sensed and/or represented (Perera et al., 2014).

By co-registering the interpreted brain responses with all related situational information, a *user model* is formed. Over time, such a model will come to reflect regularities in the users' responses and thus allow aspects of higher cognition functions to be inferred, as e.g. in Zander & Krol et al. (2016).

Continuing our example, the music that starts playing will evoke a response from the human listener—for example, they may like or dislike the song, it may distract them or evoke a pleasant memory. The listener does not intentionally produce this response; it is simply a result of them being human, and hearing the song. The response is reflected in their brain activity.
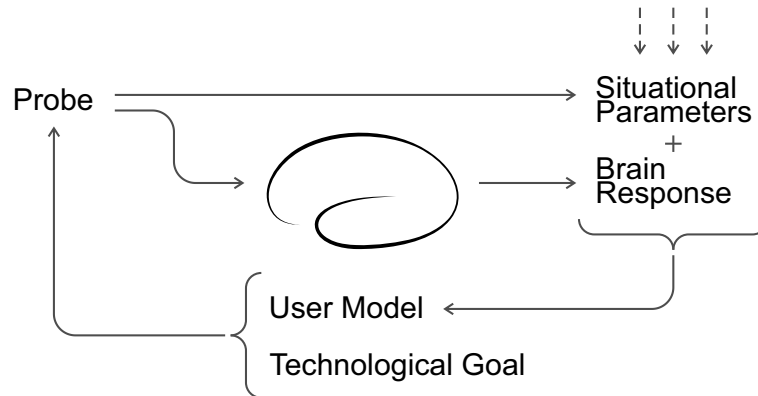
5

Figure 1: A diagrammatic overview of the cognitive probing mechanism. Based on available information, the technology initiates or manipulates a state change—a probe. This change elicits a brain response. Together with parameters describing the probe as well as, optionally, additional parameters describing the situation in which the user's response happened, the technology interprets this response and updates the model, increasing the accuracy of the information available to it for the next goal-oriented probe. At a minimum, the situation is described by the probe itself, but any other relevant parameters can be included.

The same technology that uses it to learn. The brain response induced by the probe is detected and registered in a user model alongside additional available information. The purpose of this is to learn: each response is decoded to provide at least one bit of information in order to reduce previous uncertainty or to adjust existing knowledge. This knowledge is represented in the model.

Importantly, the learner is the same technological entity that produced the probe. Combining the technological ability to interpret ongoing brain activity with the ability of that same technology to actively elicit this brain activity is what allows probes to be purposefully generated to fulfil a specific goal. An independent, external event, even if the technology interprets it and learns from it as described above, is not a cognitive probe because the technology had no influence over it.

Finishing the example, the automated, neuroadaptive DJ measures and interprets the listener's implicit brain response to the song, and learns that they do or do not like this music. It uses this information to select the next song accordingly.

Summary. Figure 1 shows a diagrammatic overview of the cognitive probing mechanism. A probe elicits a certain response in the user's brain. The neuroadaptive technology measures this response, along with relevant parameters of the state change, and co-registers these pieces of information alongside one another in a user model. The combination of all pieces of information in the model allows inferences to be drawn. Based on the user model and the current goal of the technology, the next probe is initiated. This again elicits a response, leading to an update of the user model.

To further delineate the method of cognitive probing, we mention a number of counter-examples in the discussion section of applications that are closely related to cognitive probing, but fall outside of this definition. For now, we will first turn to a number of past and present applications and experiments that do fit this definition, to

6

illustrate the method.

3   Literature Review

The proposed definition of cognitive and affective probing is deliberately generic, and does not specify, for example, for what purpose the learned information is to be used, or how inferences are to be made. Therefore, we can identify a number of past and present methods that fit this definition. Although these methods have been treated disparately before, we can see that the same terminology applies.

We may say that the history of cognitive probing starts with the *event-related potential* (ERP) technique (Luck, 2014). This was the first of a number of techniques that allowed researchers to associate the brain's neuroelectric responses with specific events. Initially, the ERP technique was used both to elucidate the neural pathways involved in sensory processing (e.g., Cobb & Dawson, 1960), and as a diagnostic tool to assess the (mal)functioning of the involved structures and processes (e.g., Chiappa, 1997). Later, it was demonstrated that ERPs also reflect aspects of cognitive processing (Walter, Cooper, Aldridge, McCallum, & Winter, 1964).

The ERP technique has been used in various applications and research endeavours. For example, Fischer et al. (1999) used a mismatch negativity paradigm (Näätänen, Paavilainen, Rinne, & Alho, 2007) to assess levels of consciousness in comatose patients, by presenting them with sentences that did or did not make semantic sense. Farwell and Donchin (1991), based on earlier work by Rosenfeld et al. (1988), presented participants with pictures of objects that did or did not belong to a crime scene in order to infer whether or not they had incriminating knowledge of it. Most ERP studies, however, including these two examples, do not employ automatic learning by the same computer that produces the stimuli, meaning they do not employ cognitive probing as such. Instead, the average ERPs are examined post hoc by the researchers. A key element that was missing was the ability to detect and analyse brain responses in real time. This had been demonstrated first in the 1970s (Vidal, 1977), but resurfaced more prominently in the 2000s with the advent of machine learning in EEG research (Ramoser, Müller-Gerking, & Pfurtscheller, 2000; Blankertz, Curio, & Müller, 2002; Lotte, Congedo, Lécuyer, Lamarche, & Arnaldi, 2007).

This development is clearly seen in reactive BCI applications (Zander & Kothe, 2011), where ERPs are elicited and used to assess correlates of human attention to allow users to voluntarily control applications. In particular, mental spellers highlight or present different letters of the alphabet on a display, in order to learn which letter was attended to (Farwell & Donchin, 1988; Treder, Schmidt, & Blankertz, 2011). Each such presentation results in an ERP. Because ERP responses to attended letters are known to be different from those to unattended ones, the system can detect this in real time and learn which letter the user intended to write. These highlights and presentations are examples of cognitive probes: they are solely initiated for the purpose of learning which letter the user wishes to communicate. Later developments added adaptive probing techniques where the sequence of highlights was adapted in real time to the available information (e.g., Lenhardt, Kaper, & Ritter, 2008; Jin et al., 2011), or where intermediate feedback elicited additional brain responses (Dal Seno, Matteucci, & Mainardi, 2010). Finally, Mladenovic et al. (2019) present an elaborate implementation of a mental speller that combines multiple adaptive strategies using a single framework based on active inference. Here, a prominent user model is used to generate the system's actions,

the user's brain responses to which in turn are used to update that user model. This is a clear example of cognitive probing in the current sense of the term.

A similar development from independent analyses of brain activity to cognitive probing as described here happened in the field of neuroergonomics (Parasuraman, 2003), where ERPs have been used for, among other things, workload detection. It has long been shown that certain features of brain responses to rare sound events seemingly reflect the availability of attentional resources (Sirevaag, Kramer, Coles, & Donchin, 1989). This led to the development of the secondary oddball task, where sounds are played repeatedly and intermixed with rare deviant sounds. Real-time evaluation of the brain responses to these tones then allow real-time inferences to be made with respect to available resources, i.e., workload. In such a context, the tones are thus cognitive probes, allowing the same technology that produces them to learn about the cognitive state of the user. This has been used in online, closed-loop contexts, where e.g. additional tasks are automated or postponed when high workload is detected in order to alleviate the load (Kohlmorgen et al., 2007). Secondary oddball tasks have also been shown to work in demanding real-world scenarios (Dehais et al., 2018).

In the previous examples, the state changes making up the probes were induced for the sole purpose of learning. However, this must not necessarily be the case: probes can also be embedded in state changes that are produced for other reasons. One example of this is the presentation of feedback in such a way as to elicit an error-related potential when the presented information is false or undesired (e.g. Dal Seno et al., 2010). Presentation of feedback is obviously an inherent part of any interactive system, its primary purpose being to inform the user. Its presentation can, however, be controlled by the computer allowing it to be used as a cognitive probe.

Similarly, Kirchner et al. (2016) effectively embedded a type of secondary oddball task in an otherwise natural human-computer interaction. During a multi-robot control task, participants had to pay attention to task-relevant messages that were presented during the interaction. The appearance of these messages evoked brain responses with properties comparable to those evoked by a dedicated secondary oddball paradigm. This allowed the computer to deduce an index of task engagement from these responses, and adapt the presentation of the messages-cum-probes accordingly. This shows that probes can be effectively hidden in natural aspects of an already-ongoing interaction.

It is even possible to have probes be the primary basis of the interaction. Zander et al. (2014); Zander, Krol, Birbaumer, and Gramann (2016) as well as Iturrate, Chavarriaga, Montesano, Minguez, and Millán (2015) demonstrated applications where this was the case: the guidance of a cursor or a robotic arm, respectively. Each movement of the cursor or arm elicited an ERP that could be automatically interpreted as reflecting the appropriateness of that movement with respect to reaching a target position. The initial movements occurred randomly. Based on the information gathered from the observer's brain response following each movement, however, a dynamic user model could be generated reflecting which behaviours were perceived positively, and which were perceived negatively. Based on this model, the movements could be steered into the apparently desired direction, towards the target. While Iturrate et al. (2015) suggest this to be used for voluntary prosthetic control, Zander & Krol et al. (2016) stress that the elicited responses themselves appear to be involuntary. In fact, their participants were unaware of having any influence over the cursor. As such, this method allowed the cursor to be controlled entirely implicitly.

This latter example presents a sliding scale with respect to the primary purpose of each cursor movement. In a process later also described by Hincks et al. (2017), the

stimuli were initially selected at random, with no other purpose than to populate the user model. As information was gathered and a user model was formed, the stimuli became increasingly part of the task completion (target reaching) mechanism while still allowing for new information to be learned.

A similar concept has been developed in the context of neuroscientific experimentation. Whereas traditional experiments are generally informed by specific hypotheses and employ correspondingly narrow experimental designs, a neuroadaptive approach to hypothesis testing would allow an experimental computer to present different stimuli or conditions in search for a specific neurophysiological response pattern (Lorenz et al., 2017). For example, in order to find which stimuli optimally activated select cortical regions, Lorenz et al. (2016) used functional magnetic resonance imaging (fMRI) to assess the participants' responses to a range of audiovisual stimuli reflecting different physical properties. Rather than presenting all stimuli in order to analyse the different responses post hoc, responses were assessed in real time. Subsequent stimuli were automatically selected based on previously gathered responses, with the aim to learn the parameters that elicit a maximum response. This closed-loop, neuroadaptive experimental design thus employs cognitive probes in order to learn how the participant responds to certain stimuli. Once the optimal stimuli have been found, the experiment can continue to gather data for further investigation.

A final example concerns the generation of a dynamic user model reflecting user interests (Wenzel, Bogojeski, & Blankertz, 2017). Here, the brain responses were not elicited by the presentation of the stimuli directly, but by the participant's eye fixation on them: the stimuli were already visible, the participants' gaze was tracked, and brain activity was analysed relative to the eyes fixating on each stimulus using so-called fixation-related potentials (FRPs; Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011). The stimuli were words of different semantic categories. By combining fixation information and an interpretation of the FRPs, the authors were able to assess which words were relevant and which were not, and ultimately, to infer which category of words the participant was interested in. This illustrates that the brain response to a specific probe (presenting the words) need not be immediate, but can be delayed. In this case, it was delayed until the participant fixated on it. This also illustrates the potential relevance of additional context information. Wenzel et al. combined the brain response with information about the participant's gaze in order to infer the semantic category of current interest.

Aside from fMRI and the EEG-based ERP technique, which featured heavily in this section, frequency-based measures can of course also be used (e.g., Ewing, Fairclough, & Gilleade, 2016; Mousavi, Koerner, Zhang, Noh, & de Sa, 2017), as well as other modalities, such as functional near-infrared spectroscopy (fNIRS) (e.g., Afergan et al., 2014; Yuksel et al., 2016).

In these examples from different areas of research, we see a convergence towards the idea of using real-time, single-trial analyses in an automated, closed-loop fashion in order to learn something about the human user or participant—i.e., towards the idea of cognitive probing presented here. While clearly sharing common methodologies, the used terminology is still disparate: authors referred to their methods, or parts thereof, as "probing" (Kane, Butler, & Simpson, 2000; Zander & Krol et al., 2016), "brain reading" or "embedded brain reading" (Kirchner, Fairclough, & Kirchner, 2019), "exploration" (Iturrate et al., 2015), "active sampling" (Lorenz et al., 2017), "active inference" (Mladenovic et al., 2019), or implied the method to be part of an "entropic BCI" (Hincks et al., 2017), of "adaptive" systems (Schultheis & Jameson, 2004; Yuksel

et al., 2016), or of a system using "brainput" (Solovey et al., 2012). This is part of the reason why we now present a unified framework and terminology.

In the next section, we discuss a number of aspects of cognitive probing that determine how implementations may differ, and aspects that are important to consider when developing systems that employ this method. In some cases, these overlap with general active learning aspects and guidelines; in others, cognitive probing must be treated separately due to its utilisation of brain data gathered directly from human beings.

4 Aspects and Considerations

Cognitive probing can be applied in different ways and across different contexts, to suit different purposes. This section lists a number of ways in which these differences can be realised. This serves both to illustrate concepts, techniques, and possibilities of the cognitive probing method, and as a list of considerations and trade-offs for researchers and developers.

Furthermore, the cognitive probing method raises a variety of ethical issues that we will briefly introduce and discuss in this section. The basic concern centres around the implications of probing someone's brain processes in order to gather cognitive or affective information about that person. A responsible use of such technology requires, at the very least, reflections on the topics of privacy and consent, and responsibility for actions mediated by the technology.

4.1 Learning Strategies

Probes are designed in order to obtain information. Two central questions are thus: What information is to be obtained, and how is this best achieved?

Ultimately, the information that can be obtained of course depends on the cognitive and affective states that can be detected and the reliability of these assessments. Here, traditional neuroscientific considerations apply. See e.g. Brouwer, Zander, van Erp, Korteling, and Bronkhorst (2015) and Gerjets, Walter, Rosenstiel, Bogdan, and Zander (2014) for recommendations.

There are different possible strategies to learn from cumulative samples and to choose subsequent probes. In the previous section, Lorenz et al. (2016) and Mladenovic et al. (2019) used Bayesian inference and optimisation (Snoek, Larochelle, & Adams, 2012), whereas Iturrate et al. (2015) used a form of reinforcement learning (Sutton & Barto, 1998).

With its abstract similarity to active learning, the cognitive probing method can utilise many of the so-called query strategies of active learning (Settles, 2009). As illustrated in Figure 1, the system can select a specific state change based on the knowledge currently available in its model, in order to update that model (i.e., to fulfil its goal of learning). It can thus use available information in order to determine which change will be initiated, and to what end.

One strategy to that effect that Settles (2009) mentions is *uncertainty sampling*: the system can initiate changes along those dimensions about which it has the least knowledge. For example, if the system has no information pertaining to a particular condition, it could deliberately produce this condition in order to probe the user's response to it.

Other strategies can similarly be applied, notably *expected model change* (obtaining information that would most change the model) and *expected error reduction* (to reduce the error in the model, or in the system's behaviour). A system may initialise using

random variables and based on these initial samples, move to specific strategies to find more suitable values (Hincks et al., 2017).

4.2 Repeating Probes

Most likely, it will be necessary to probe multiple times in order to draw reliable conclusions. Given imperfect measurement or inference methods, a trade-off needs to be considered between the number of probes and the accuracy of the gathered information.

With respect to obtaining a specific piece of information, a sequence of probes can be presented in different ways. One way is to present a precisely defined state change a pre-determined number of times (e.g., once, or ten times). The optimal number of repetitions has, for example, been investigated for P300 spellers (e.g., Jin et al., 2011). Another method is to use a confidence criterion, where the state change is repeated until a certain level of confidence has been reached (i.e., the error reduction strategy combined with dynamic stopping; e.g., Schreuder et al., 2013). Finally, rather than repeating the same probe, the probe itself could be dynamically changed during the sequence, adaptively homing in on increasingly specific information of interest as e.g. in Lorenz et al. (2017).

4.3 The Roles of Time

Time is relevant in a number of ways. For one, time itself may be one of the relevant situational factors to be included in a model: the absolute time of day (e.g., to learn patterns that are different in the evening as compared to the morning), or the time relative to a previous event.

Also the time needed for the system to measure and process the brain response can have practical importance. Sometimes the measurement and data analysis might take longer than what is useful for a particularly time-critical application.

Furthermore, the duration of a probe's effect needs to be considered. In the examples given in this paper so far, all relevant brain responses to the probe were near-immediate. However, this need not always be the case. Depending on the application, a distinction can be made between up to four time points or periods: the time that the probe is generated (i.e., the technological state is changed); the time that the probe is perceived; the period of time during which the probe's influence remains apparent (i.e., during which relevant brain responses can be gathered); and the period of time during which the gathered information remains valid.

The first two time points are usually the same, but not necessarily. It is possible for a probe to be noticed only after it has been produced, as for example in the fixation-related study mentioned above (Wenzel et al., 2017). It may also be possible for a state change not to be consciously noticed at all, while still having an effect (Negri, Gamberini, & Cutini, 2014). The third point covers a time period that varies depending on factors such as habituation, and the time it takes for a meaningful response to arise—see also section 4.7 for the difference between impulsive and reflective responses. The final point concerns the "half-life", so to say, of any gathered information. Is it generally valid and can it remain in the model indefinitely, or may the user's responses change over time such that the learned information becomes obsolete? In the latter case, the probes may need to be repeated at certain intervals.

4.4 A Probe's Purpose

The presented definition requires that probing must be done in order to learn. This does not mean, however, that learning is necessarily the only objective: the system may simultaneously have other goals.

Figure 1 illustrates three effects of a probe: its effect on the environment (at a minimum, this is simply the occurrence of the probe itself), its effect on the brain (the brain response elicited by the probe), and ultimately, the adjustment of the model. The goal of "learning" refers to this last effect. However, cognitive probes can be selected to simultaneously fulfil further strategic goals. In particular the second consequence, the user's response, can be part of the goal-driven behavioural strategy of the system.

The system may aim to induce or sustain desirable psychophysiological user states. If certain probes have been learned to lead to positive responses, the system may aim to induce or sustain the user's positive state by repeating such probes. This was the case for the implicit cursor control experiment (Zander & Krol et al., 2016): each probe served both to further the system's knowledge of the preferred direction, and to steer the cursor into that direction.

Note that goals need not remain static over time: for example, an initial phase of mere information-gathering can be followed by different goals as decided based on the gathered information, or goals may change based on changing situational aspects.

Furthermore, even when learning is the main objective to employ cognitive probing itself, the purpose for which information needs to be learned is left entirely open.

In other words, it can be said that neuroadaptive systems have their own agenda (Fairclough, 2017). An important factor to consider is whether or not, or to what extent, this agenda is in line with the user's own goals, discussed in more detail further below.

4.5 Interference and Distraction

The fact that cognitive probes are intended to elicit a response means that they must influence the user in one way or another. As mentioned above, this influence can be used by the system for a higher goal, such as the promotion of a desired mental state. At other times, this influence may be an unwanted side-effect. The effects may be distracting or otherwise bothersome, in particular when the system does not yet have a sufficiently accurate model, or when probes are selected using a seemingly incomprehensible strategy. Potentially, this could even influence their mental state to a point of no longer being useful as implicit input.

In utilising cognitive probes, it must be considered to what degree these probes may interfere with the user's natural or intended behaviour. An important factor here is how well the probes can be embedded into the natural context.

For example, the secondary oddball tasks mentioned earlier used clearly perceptible, overt auditory probes (Schultheis & Jameson, 2004; Kohlmorgen et al., 2007). These were required in order to learn of the user's attentional resources, but were otherwise irrelevant to the user. These are inherently distracting stimuli (Parmentier, Elsley, Andrés, & Barceló, 2011). On the other hand, the example of task-relevant messages that were co-opted as probes (Kirchner et al., 2016) shows that they may have been distracting, but not because they were probes—their use as probes did not change their nature as part of the interaction. The messages themselves were clearly perceptible, but their function as probe was not noticeable at all. This can also change over time. During the cursor and robot control examples (Zander & Krol et al., 2016; Iturrate et al., 2015),

the initial actions were purely information-gathering probes, but later actions, based on an increasingly accurate model, were increasingly compatible with the user's goals, and their probing nature became secondary to the goal of steering the cursor to the target.

Probes that are embedded in naturally-occurring events are less distracting, but must wait until those events happen. Also note that, the better probes are "hidden", the more care must be taken not to violate the user's privacy, discussed in section 4.9. Alternatively, a user can be exposed to a calibration phase of clearly perceivable and attended, but seemingly unproductive probes until the model is sufficiently accurate for the probes to be helpful. In other cases, of course, e.g. where the process always has the full attention of the user, these deliberations may not apply.

4.6   Representing Multiple Dimensions

Theoretically, the model that contains the gathered information can be of any level of complexity or dimensionality. Not only can any number of probe-response events be taken into account, any number of additional context measurements can be registered along with that response.

Most applications can be broken down into discrete dimensions: a light, for example, may be determined by its hue and its intensity. Cognitive probes could be used to gather a person's current preference with respect to these two dimensions. However, such preferences may be dependent on further dimensions: the time of day, their current activity in the room, whether or not there are guests... In this sense, the full complexity of the real world has an indeterminate number of dimensions.

Each probe can be seen as a single point in a large state space covering all dimensions in the model. Section 4.1 already covered some strategies to explore this space. An additional consideration is the desired accuracy on each individual dimension. Disentangling the contributions of individual dimensions will require a large amount of probes, and until such time, no meaningful inferences can be drawn.

Alternatively, the complexity may be increased step-wise, with additional dimensions added only in due time, after prior models have been properly calibrated. We may also note that not all information must necessarily be obtained through cognitive probing. One could pre-populate a user model with expected values, and have the system select cognitive probes using expected model change strategies to quickly adapt the model to the momentary preferences of the individual user. Similarly, multiple archetypal user models might be pre-populated, e.g. using statistical inference based on large amounts of data, and cognitive probes can be used to select between them for any given moment or person.

4.7   Impulsive versus Reflective Responses

Because cognitive probing relies on externally induced brain responses, rather than self-initiated responses, it is possible that these responses are the result of shallow (impulsive, intuitive, reflexive) processing, subject to heuristics and biases (Kahneman, 2011). In that case, it is important to know to what extent the induced brain responses (would) correspond to explicitly given responses.

Refer, for example, to Frankfurt (1971)'s distinction between first- and second-order volitions, where first-order volitions reflect first impulse desires (e.g., wanting to smoke a cigarette), while second-order volitions (wanting to stop smoking) reflect our self-control, reason-based cognition. It is conceivable that a person, given the time to deliberate

and formulate an explicit response to a probe, would come to a different one than the one induced and inferred by the technology. As such, cognitive probes could result in information being implicitly communicated by the user that the user does not or would not authorise.

Cognitive probing does not prescribe using either impulsive or reflective brain response. It is, however, an important distinction for researchers and developers to be aware of, as it concerns the type of information that can be learned using cognitive probes in different contexts, and the extent to which this information and subsequent adaptations may be accepted by the user. Furthermore, this can have profound effects on responsibility assessment, discussed next.

4.8 Outcome Responsibility

Who or what is responsible for an outcome that is based on cognitive probing? As mention in section 4.4, the technology acts upon the information it gathers by modulating future probes in order to learn, but it can also use that information to execute different, additional actions.

When technology contributes autonomously to the performance of a task a user is engaged in, it may become difficult to separate the two interacting agents. How would the user (or others) evaluate the respective contributions of human and technology? By whose authority is an action executed? To whom would "ownership" of the action be ascribed, or responsibility for its outcome? Even while performing an action, the users themselves might be uncertain about being the (only) agent of an action (Haselager, 2013), with systems that make autonomous decisions additionally decreasing the users' own autonomy (Friedrich, Racine, Steinert, Pömsl, & Jox, 2018).

Because cognitive probing relies on externally induced brain responses, one could argue that behavioural outcomes based on this information would fall within the reflexive or spontaneous behaviour class, or at least that the underlying thought process was not finished (see also section 4.7). Users could attempt to disavow such behaviours as not being congruent with their own reason-based ideas about how they ought to act. Users who do not agree with (the use that is being made of) their own brain responses may argue that responsibility lies elsewhere, since their input to the system bypassed their conscious awareness and control.

Moreover, probing technology might lead to a situation where users could reasonably feel that in cases of continuous deepening probes (probes provided on the basis of cognitive information gathered from previous probes), their capacity to engage in autonomous reasoning and decision making may be compromised. In essence, one needs to consider the implications of "nudging" technology being applied at the level of brain processes, opening up further possibilities for user manipulation (see also section 4.4).

There are different types, or ways of claiming, ownership of action. Under normal circumstances we accept our actions to be ours, irrespective of whether they were of a reflexive nature, spontaneous first impulses, or following deliberation. However, when it comes to assessing the moral consequences of such behaviours, we may differentiate between them. The assessment of my responsibility, e.g. my guilt, based on my behaviour that led to damage, e.g. a broken glass, may differ depending on whether my behaviour was a mere reflex, a spontaneous movement, or a consequence of a reasoned judgement.

Under such circumstances, it seems possible that users will "blame" the technology in cases with negative outcomes, but praise themselves in cases of success.

The effect of cognitive probing on assessing the responsibility for consequences in

legal or financial contexts will be far from straightforward. Presumably, many of such consequences could be dealt with by using liability waivers, although this could affect the willingness of users to accept cognitive probe technology in realistic contexts (with realistic consequences). Also note that procedures and criteria for accepting or attributing responsibility may diverge significantly between legal, and moral domains. It is important that researchers and designers of systems with direct access to brain activity and the ability to act autonomously ensure that they act with the user's permission and in accordance with the user's wishes (Fairclough, 2009; Mecacci & Haselager, 2019b).

4.9 Privacy and Consent

The interference discussed in section 4.5 referred to the immediate intrusion that individual probes could pose into a user's state of mind. The method of cognitive probing in general, however, could also be understood to intrude more fundamentally into a person's privacy, as it elicits and utilises brain responses that are beyond the direct control of the user. As such, cognitive probing allows technology to autonomously gather or "extract" information from the user, touching upon the privacy of thought. As discussed above, this information gathering can potentially take place unbeknownst to the user—and thus, potentially, without consent.

Therefore, care must be taken that the data is handled with proper care for the user's privacy and the security and confidentiality of all gathered data (Fairclough, 2014). Brain data in particular, being more closely connected to the privacy of "inner" thoughts and feelings, may be all the more sensitive. Moreover, by utilising automatic brain responses, the information acquired may reflect mental features that, to some extent, remain out of reach of a person's self-control, meaning data may be gathered that would not have been voluntarily provided, or would not even be known by the users themselves.

Hence, users of cognitive probing applications should be given full transparency: they should be informed clearly and unequivocally about the nature, methodology, and aims of the technique (see also Ienca, Haselager, & Emanuel, 2018). Specifically for cognitive probing technology, at the very least, they should be informed that *a)* cognitive probes induce brain responses that the users normally cannot suppress, *b)* such probes and responses can and will be used to gather information about the user's cognitive and affective states, and that *c)* this information, in addition to contextual information that is also being gathered, will be used to provide more and possibly better directed probes in order to extract even more information.

Haselager and Mecacci (2018; see also Mecacci & Haselager, 2019a) discuss applications of brain measurement aimed at brain reading, i.e., decoding mental content. This discussion applies to cognitive probing as well, as both methods capture cognitive information. They suggest that two factors are of vital importance with respect to potential abuse: concealability, and enforceability.

Concealability indicates the extent to which a user can be left unaware of the actual goals behind the particular application, e.g. the nature, amount, or detail of the cognitive or affective states that are being explored. Note that the mere knowledge that one's brain states are being monitored and probed does not imply an awareness of which aspects of cognition are being investigated.

Enforceability is related to the extent to which the application is resilient to a subject's voluntary attempt to disrupt the process. In the current context, this implies that once subjects are participating in the process, the more "automatic" their responses to

a probe are, the more enforceable the information gathering will be.

Consideration of these two factors leads to the further requirement that *d)* users will be informed about the exact nature, amount, and detail of the information about cognitive and affective states explored, with in addition *e)* the purposes for which this information is being gathered.

Given the sensitive nature of such gathered information, *f)* procedures and guarantees regarding the ownership of data should be installed ensuring that users of cognitive probing applications remain in control of the data acquired, its storage, and its future use (Fairclough, 2014; Yuste et al., 2017). As one part in safeguarding the users from the above issues, *g)* the user must have access to a comprehensive and meaningful representation of all information in the model, and insight into the logic of the system's probes and model-based behaviour, in line with the recent General Data Protection Regulation (GDPR) of the European Union (European Parliament and Council of the European Union, 2016).

However, it is not only brain-related information, or its interpretation in mental terms, but also the ensuing operations of the technology that might constitute threats to (mental) privacy of its users. Probe-based technological state changes may lead to publicly noticeable outcomes that, even when in agreement with a user's genuine preference, may be socially problematic or undesirable. At the very least, therefore, *h)* users should be aware that aspects of their mental states may be revealed to outsiders in the form of probes or other adaptations.

Of course, we do not wish to claim that this list of requirements is complete or formulated precisely enough. Moreover, the stringency with which they should be applied will depend considerably on the context and purposes of the applications. However, cognitive probing does pose serious threats to privacy and consent, and this section specifies a number of concerns which researchers and developers are called on to consider and discuss.

5   Discussion

Cognitive probing combines the ability of technology to interpret ongoing measures of arbitrary brain activity, with the ability of that same technology to actively and purposefully elicit cognitive and affective responses from its users. This is used to learn and infer information from these responses. Based on the resulting knowledge, additional, increasingly targeted responses can be elicited.

Looking at the different commercial, clinical, and experimental approaches and applications mentioned in section 3, we believe the concept of cognitive probing as introduced here describes the fundamental similarity between a number of previously separate endeavours. Some early incarnations date back to decades ago, but along with an increasing ability to automatically assess and interpret mental states has come an increasing number of recent, novel examples that fit this same definition of cognitive probing. Some stem from strictly controlled laboratory environments, but given the current commercialisation of neuroadaptive technology, a growing number of examples are now also preparing for general use. Going forward, it is important to recognise the fundamental similarities shared by these different applications, especially since the specific configuration that defines cognitive probing gives rise to a number of serious ethical considerations.

To further delineate the concept of a cognitive probe, we now consider some counter-

16

examples that deviate from the definition in specific ways. Without implying that less care should be taken in these cases, the following examples do fall outside of the specific scope of cognitive probing.

As one example, we may consider a case where all other requirements are met but the response-eliciting event in question was not influenced by the system at all, i.e. it was not initiated or manipulated by it. For example, instead of a radio being switched on, a colleague starts whistling a tune. A context-aware system may still register this, identify it as a noteworthy event, record the listener's brain response to it, and infer whether or not this tune is distracting. Since the colleague acts independently, it cannot be said that their behaviour is a probe from the system.

When the state change is not independent from the technology, a crucial question becomes what exactly counts as being "initiated" or "manipulated" by it. A text message that you receive on your phone was initiated by the sender, but the notification you receive of it may not have been, depending on how your phone handles notifications. If the phone is programmed to always automatically immediately notify you of all messages that come in, the phone has no control over them and they should be seen as being initiated by the sender of those messages. If your phone does have policies in place that allow it to manipulate the notifications, they can be used as cognitive probes.

Of course a neuroadaptive system may use any and all information available to it, including all responses to external or otherwise non-probe events. Doing so, however, would not necessarily constitute cognitive probing, but more generally the collection, selectively or indiscriminately, of event-response samples.

As another counter-example, a system may have access to a measure of its user's brain activity and initiate a state change to get a specific response—but without learning from it. Instead, it could be initiated for the sake of the response itself. A car detecting a state of drowsiness in its driver may decide to sound an alarm in order to awaken the driver. This would be a state change initiated by the car designed to elicit a specific response from its user, but it would not be a cognitive probe as the loop is not closed and no information is retained.

Other brain responses may not contain any cognitive or affective information per se. BCI systems relying on steady-state visually evoked potentials (SSVEPs; Middendorf, McMillan, Calhoun, & Jones, 2000) for example, make use of the fact that when a human is presented with a visual stimulus that is flashing at a specific frequency, their visual cortex will exhibit neuro-electric activity at that same frequency (and its harmonics). When different stimuli thus flash at different frequencies, the dominant frequency in the visual cortex can reveal what stimulus a user is looking at. Thus, SSVEP-based BCI systems present system-controlled state changes (the flashing of the visual stimuli), monitor the user's brain response to these state changes, and infer from the found frequency which letter the user was looking at (i.e., wanted to select). This is not cognitive probing, however, as the brain response in question is not related to the user's cognition. The activity in the visual cortex is a direct result of the sensory input; there is no relevant top-down cognitive influence. This separates this method from e.g. attention-based mental spellers, where cognition does influence the measured response. Furthermore, these applications tend to be one-shot approaches with no model learning.

Finally, not all closed-loop neuroadaptive systems (Krol et al., 2018) use cognitive probing. Adaptive automation, for example, adjusts automation levels based on a measurement of workload in order to keep this workload within a desired range (Byrne & Parasuraman, 1996). This principle has also been used to make sure students sustain an adequate level of engagement (Yuksel et al., 2016). Here, when a measure of load

surpasses a certain threshold, the technology changes the conditions by e.g. increasing or decreasing automation, or presenting material at an appropriate level of difficulty. These changes are designed to influence the user and maintain a desired psychological state. When, following this, the measure again passes the threshold, further adjustments are made. This method has many aspects in common with cognitive probing, but these examples do not employ a model where responses to specific conditions are saved. These systems demonstrate the idea that directed adaptations based on an ongoing evaluation of a user's state can influence that same state, but they stop short of using that influence and evaluation to automatically learn from the effects of such adaptations—although they could easily be extended to do so.

As mentioned before, examples or forerunners of cognitive probing can be found going back decades, with similar developments happening in parallel. This has resulted in essentially equivalent approaches being developed independently, and being given separate names. In our proposed unification, we elected for the label "cognitive probing", or "affective probing", as *probe* was one of the first terms that was and still is being used for computer-presented stimuli aimed at eliciting specific brain responses (Kane et al., 2000; Espie, 2007; Lawrence et al., 2014). We must note that the same term has found some use also in the fields of questionnaire design, where it refers to follow-up questions targeting specific previous statements (Beatty & Willis, 2007), and in psychotherapy, where a cognitive probe may be such a question as "What are you thinking right now?" (Beck, 1991). The semantic overlap here, we believe, does not interfere with the more specific meaning in the context of neuroadaptive technology, but rather demonstrates that this term readily reflects important parts of the intended meaning.

Cognitive probing represents only a specific method in the larger field of neuroadaptive technology, but it is a powerful one. It allows neuroadaptive technology to escape the confines of a single cybernetic loop (Pope, Bogart, & Bartolome, 1995), and, in effect, allows it to autonomously pose questions to the user, obtaining an implicit answer directly from the user's elicited brain activity. Why it asks those questions and what it subsequently does with the answers—the options are innumerable.

6   Outlook

Traditionally, machines only respond to their operators' explicit commands. Using passive brain-computer interfaces and further forms of physiological computing, it is possible to incorporate automatic interpretations of user states as secondary input modalities. With cognitive probing, it is possible to not merely rely on the naturally available information, but to actively and autonomously gather information that was not otherwise available from the user—potentially even unbeknownst to them.

The example of a neuroadaptive book, described by Zander & Krol et al. (2016), provides an interesting hypothetical scenario for cognitive and affective probing:

> While reading, the reader would interpret the story as it unfolded, thus automatically responding to events with a detectable affective state. Based on what the reader apparently finds enjoyable, a neuroadaptive system could then change the content of subsequent pages. With a sequence of such adaptations, the story is gradually steered in the reader's preferred direction. However, the reader would not actively be directing the story, and would not even need to be aware of the system's existence. (Zander & Krol et al., 2016)

Interactive narratives using neurophysiological input have been presented before (Gilroy et al., 2013), but none where the events unfolding on the pages serve both as the content of the book, and as probes. They are neither distracting nor intrusive, as they are perfectly embedded into, in fact *are*, the story itself. The first few chapters may be pre-determined, designed to learn the reader's attitudes towards, say, the main characters. Should, however, this not suffice, the book could decide to present the reader with another chapter focusing more thoroughly on the one character it is least certain of (i.e., uncertainty sampling).

In this example, the complexity of the model depends largely on the amount of possible paths the book will have. It could be merely a decision between two outcomes, or more complex information can be gathered to allow more detailed plot developments and twists. In an extreme case, language models may be used to generate new paragraphs and adaptive probes for an intricate model and potentially endless storyline (Radford et al., 2019). The purpose of the book as a whole could be, for example, to provide a pleasant experience (sustain a positive psychological state), to induce as many surprising plot twists as possible, or even to reveal deep, personal values.

Either way, the final contents of the book will at least partially reflect the reader's brain responses. The example thus also makes it clear how the very behaviour and outcome of a neuroadaptive system may reveal sensitive information, as discussed in section 4.9. While cognitive probing-based neuroadaptive systems allow for a unique level of personalisation, would the reader wish to reveal to others that, for example, evil prevailed in their version of the book?

An illustrative use case for non-embedded cognitive probing may be found with online retailers. After selecting a category, a number of stock items may be presented and the user's reaction probed, until a recommendation is made tailored to the user's inferred preferences.

Here, it is obvious to the user that they are being probed, and for what purpose. In the case of the neuroadaptive book, it may be altogether unclear which parts of the rich prose exactly comprise the probes and what information they attempt to gather. We emphasise the ethical issues mentioned in section 4, in particular concealability and transparency.

In theory, a neuroadaptive system can be imagined in a ubiquitous computing approach, i.e., with access to a large amount of context information concerning the user's many everyday interactions with their environment. If even a single cognitive user state—positive versus negative, satisfaction versus dissatisfaction—could be accurately inferred and co-registered with these many interactions and other context information, a user model could be generated that, over time, could contain a very detailed description of this user's habits and preferences. Cognitive probes may be used to expand this user model with increasing efficiency. A system that continuously learns from our implicit responses to its directed probes, and thus generates a rich understanding of our preferences and intentions, may even be in a position to exhibit a form of intelligence and empathy, and support us in a way not currently possible by robots or other machines.

From a different angle, such a system might also be used as a tool to influence the cognition and the mindset of a person in an interplay between the probes' ability to obtain information and its ability to produce specific responses. Imagine a (social) media platform with access to brain activity, where each presented piece of media could constitute a probe. Even inferences of seemingly simple information—for example, colour preferences in order to adjust the interface's colour scheme—could be used in potentially dangerous ways: politicians of different parties could be deliberately presented with dif-

ferent background colours, biasing the user to dislike the politician presented outside of the preferred colour scheme under the guise of advertisement personalisation. Cognitive probes may also be capable of assessing political preferences directly. With cognitive probing's ability to obtain increasingly specific information and target specific cognitive or affective functions, even outside of the user's awareness, it may simultaneously produce increasingly subtle methods of manipulating the user's mindset. We mention this example to again highlight the importance of consent and transparency.

A number of aspects and considerations mentioned in section 4 remain open to investigation. Can we distinguish between impulsive and reflective responses? To what extent are they different, neurophysiologically and behaviourally? Which ones could be targeted using what probes? How do users judge agency and responsibility in different cognitive probing cases? What influences this? What are potential consumers' attitudes towards cognitive probing from e.g. a privacy standpoint? Answers to such questions will help determine what can and cannot, and what should or should not be achieved using cognitive and affective probes.

Going forward, provided that researchers and developers properly discuss and address these and other serious ethical concerns, we believe that cognitive probing can help make technology more intelligent, more interactive, and more adaptive to their users' needs and preferences.

7 Acknowledgements

References

Afergan, D., Peck, E. M., Solovey, E. T., Jenkins, A., Hincks, S. W., Brown, E. T., ... Jacob, R. J. (2014). Dynamic difficulty using brain metrics of workload. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 3797–3806). New York, NY, USA: ACM. doi: 10.1145/2556288.2557230

Beatty, P. C., & Willis, G. B. (2007). Research Synthesis: The Practice of Cognitive Interviewing. *Public Opinion Quarterly, 71*(2), 287–311. doi: 10.1093/poq/nfm006

Beck, A. T. (1991). Cognitive therapy: a 30-year retrospective. *American Psychologist, 46*(4), 368–375.

Blankertz, B., Curio, G., & Müller, K.-R. (2002). Classifying single trial EEG: Towards brain computer interfacing. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems 14* (pp. 157–164). Cambridge, MA, USA: MIT Press.

Brouwer, A.-M., Zander, T. O., van Erp, J. B. F., Korteling, J. E., & Bronkhorst, A. W. (2015). Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Frontiers in Neuroscience, 9*, 136.

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology, 42*(3), 249–268. doi: 10.1016/0301-0511(95)05161-9

Chiappa, K. H. (1997). *Evoked potentials in clinical medicine* (Third ed.). Philadelphia, PA: Lippincott-Raven Publishers.

Cobb, W. A., & Dawson, G. D. (1960). The latency and form in man of the occipital potentials evoked by bright flashes. *The Journal of Physiology, 152*(1), 108–121. doi: 10.1113/jphysiol.1960.sp006474

Dal Seno, B., Matteucci, M., & Mainardi, L. (2010). Online detection of P300 and error potentials in a BCI speller. *Computational Intelligence and Neuroscience, 2010*, 11:1–11:5. doi: 10.1155/2010/307254

Dehais, F., Dupres, A., Di Flumeri, G., Verdiere, K., Borghini, G., Babiloni, F., & Roy, R. N. (2018). Monitoring pilot's cognitive fatigue with engagement features in simulated and actual flight conditions using an hybrid fNIRS-EEG passive BCI. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 544–549). doi: 10.1109/SMC.2018.00102

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General, 140*(4), 552–572. doi: 10.1037/a0023885

Donchin, E. (1981). Surprise!. . . Surprise? *Psychophysiology, 18*(5), 493–513.

Espie, C. A. (2007). Understanding insomnia through cognitive modelling. *Sleep Medicine, 8*, S3–S8. doi: 10.1016/S1389-9457(08)70002-9

European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union, L 119*, 1–88.

Ewing, K. C., Fairclough, S. H., & Gilleade, K. (2016). Evaluation of an adaptive game that uses EEG measures validated during the design process as inputs to a biocybernetic loop. *Frontiers in Human Neuroscience, 10*, 223. doi: 10.3389/fnhum.2016.00223

Fairclough, S. H. (2009). Fundamentals of physiological computing. *Interacting with computers, 21*(1), 133–145. doi: 10.1016/j.intcom.2008.10.011

Fairclough, S. H. (2014). Physiological data must remain confidential. *Nature, 505*, 263. doi: 10.1038/505263a

Fairclough, S. H. (2017). Physiological computing and intelligent adaptation. In M. Jeon (Ed.), *Emotions and affect in human factors and human-computer interaction* (pp. 539–556). San Diego, CA, USA: Academic Press. doi: 10.1016/B978-0-12-801851-4.00020-3

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: a tutorial. *Biological Psychology, 51*(2-3), 87–107.

Farwell, L. A., & Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology, 70*(6), 510–523.

Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy ("lie detection") with event-related brain potentials. *Psychophysiology, 28*(5), 531–547. doi: 10.1111/j.1469-8986.1991.tb01990.x

Fischer, C., Morlet, D., Bouchet, P., Luaute, J., Jourdan, C., & Salord, F. (1999). Mismatch negativity and late auditory evoked potentials in comatose patients. *Clinical Neurophysiology, 110*(9), 1601–1610. doi: 10.1016/S1388-2457(99)00131-5

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy, 68*(1), 5–20.

Friedrich, O., Racine, E., Steinert, S., Pömsl, J., & Jox, R. J. (2018). An analysis of the impact of brain-computer interfaces on autonomy. *Neuroethics*. doi: 10.1007/s12152-018-9364-9

Gates, C., Subramanian, S., & Gutwin, C. (2006). DJs' perspectives on interaction and awareness in nightclubs. In *Proceedings of the 6th Conference on Designing Interactive Systems* (pp. 70–79). New York, NY, USA: ACM. doi: 10.1145/1142405.1142418

Gerjets, P., Walter, C., Rosenstiel, W., Bogdan, M., & Zander, T. O. (2014). Cognitive state monitoring and the design of adaptive instruction in digital environments: lessons learned from cognitive workload assessment using a passive brain-computer interface approach. *Frontiers in Neuroscience, 8*, 385. doi: 10.3389/fnins.2014.00385

Gilroy, S. W., Porteous, J., Charles, F., Cavazza, M., Soreq, E., Raz, G., . . . Hendler, T. (2013). A brain-computer interface to a plan-based narrative. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence* (pp. 1997–2005). Palo Alto, CA, USA: AAAI Press.

Haselager, P. (2013). Did I do that? Brain-computer interfacing and the sense of agency. *Minds and Machines, 23*(3), 405–418. doi: 10.1007/s11023-012-9298-7

Haselager, P., & Mecacci, G. (2018). Is brain reading mind reading? In B. Donnelly-Lazarov, D. Patterson, & P. Raynor (Eds.), *Neurolaw and responsibility for action: Concepts, crimes and courts* (pp. 182–191). Cambridge, UK: Cambridge University Press. doi: 10.1017/9781108553339

Haynes, J.-D. (2011). Brain reading: Decoding mental states from brain activity in human. In J. Illes & B. J. Sahakian (Eds.), *Oxford handbook of neuroethics* (pp. 3–13). New York, NY, USA: Oxford University Press. doi: 10.1093/oxfordhb/9780199570706.013.0013

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences, 95*(3), 781–787.

Hincks, S. W., Bratt, S., Poudel, S., Phoha, V. V., Jacob, R. J. K., Dennett, D. C., & Hirshfield, L. M. (2017). Entropic brain-computer interfaces - using fNIRS and EEG to measure attentional states in a bayesian framework. In *Proceedings of the 4th International Conference on Physiological Computing Systems (PhyCS)* (pp. 23–34).

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature, 532*, 453–458. doi: 10.1038/nature17637

Ienca, M., Haselager, P., & Emanuel, E. J. (2018). Brain leaks and consumer neurotechnology. *Nature Biotechnology, 36*, 805–810. doi: 10.1038/nbt.4240

Iturrate, I., Chavarriaga, R., Montesano, L., Minguez, J., & Millán, J. d. R. (2015). Teaching brain-machine interfaces as an alternative paradigm to neuroprosthetics control. *Scientific Reports, 5*, 13893. doi: 10.1038/srep13893

Janssen, J. H., Tacken, P., de Vries, G.-J., van den Broek, E. L., Westerink, J. H. D. M., Haselager, P., & IJsselsteijn, W. A. (2013). Machines outperform laypersons in recognizing emotions elicited by autobiographical recollection. *Human-Computer Interaction, 28*(6), 479–517. doi: 10.1080/07370024.2012.755421

Jin, J., Allison, B. Z., Sellers, E. W., Brunner, C., Horki, P., Wang, X., & Neuper, C. (2011). An adaptive P300-based control system. *Journal of Neural Engineering, 8*(3), 036006. doi: 10.1088/1741-2560/8/3/036006

Kahneman, D. (2011). *Thinking, fast and slow.* New York, NY, USA: Farrar, Straus and Giroux.

Kane, N. M., Butler, S. R., & Simpson, T. (2000). Coma outcome prediction using

event-related potentials: P3 and mismatch negativity. *Audiology and Neurotology*, *5*(3–4), 186–191. doi: 10.1159/000013879

Kirchner, E. A., Fairclough, S. H., & Kirchner, F. (2019). The handbook of multimodal-multisensor interfaces. In S. Oviatt, B. Schuller, P. R. Cohen, D. Sonntag, G. Potamianos, & A. Krüger (Eds.), (pp. 523–576). New York, NY, USA: AMC and Morgan & Claypool. doi: 10.1145/3233795.3233810

Kirchner, E. A., Kim, S. K., Tabie, M., Wöhrle, H., Maurus, M., & Kirchner, F. (2016). An intelligent man-machine interface—multi-robot control adapted for task engagement based on single-trial detectability of P300. *Frontiers in Human Neuroscience*, *10*, 291. doi: 10.3389/fnhum.2016.00291

Kleinsmith, A., Bianchi-Berthouze, N., & Steed, A. (2011). Automatic recognition of non-acted affective postures. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *41*(4), 1027–1038. doi: 10.1109/TSMCB.2010.2103557

Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Curio, G., Hagemann, K., . . . Kincses, W. (2007). Improving human performance in a real operating environment through real-time mental workload detection. In G. Dornhege (Ed.), *Toward Brain-Computer Interfacing* (pp. 409–422). Cambridge, MA, USA: MIT Press.

Krol, L. R., Andreessen, L. M., & Zander, T. O. (2018). Passive Brain-Computer Interfaces: A Perspective on Increased Interactivity. In C. S. Nam, A. Nijholt, & F. Lotte (Eds.), *Brain-Computer Interfaces Handbook: Technological and Theoretical Advances* (pp. 69–86). Boca Raton, FL, USA: CRC Press.

Krol, L. R., Freytag, S.-C., Fleck, M., Gramann, K., & Zander, T. O. (2016). A task-independent workload classifier for neuroadaptive technology: Preliminary data. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 003171–003174). doi: 10.1109/SMC.2016.7844722

Krol, L. R., & Zander, T. O. (2018). Towards a conceptual framework for cognitive probing. In J. Ham, A. Spagnolli, B. Blankertz, L. Gamberini, & G. Jacucci (Eds.), *Symbiotic interaction* (pp. 74–78). Cham, Switzerland: Springer International Publishing. doi: 10.1007/978-3-319-91593-7_8

Lawrence, E. J., Su, L., Barker, G. J., Medford, N., Dalton, J., Williams, S. C., . . . David, A. S. (2014). Self-regulation of the anterior insula: Reinforcement learning using real-time fmri neurofeedback. *NeuroImage*, *88*, 113–124. doi: 10.1016/j.neuroimage.2013.10.069

Lenhardt, A., Kaper, M., & Ritter, H. J. (2008). An adaptive P300-based online brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *16*(2), 121–130. doi: 10.1109/TNSRE.2007.912816

Lorenz, R., Hampshire, A., & Leech, R. (2017). Neuroadaptive bayesian optimization and hypothesis testing. *Trends in Cognitive Sciences*, *21*(3), 155–167. doi: 10.1016/j.tics.2017.01.006

Lorenz, R., Monti, R. P., Violante, I. R., Anagnostopoulos, C., Faisal, A. A., Montana, G., & Leech, R. (2016). The automatic neuroscientist: A framework for optimizing experimental design with closed-loop real-time fMRI. *NeuroImage*, *129*, 320–334. doi: 10.1016/j.neuroimage.2016.01.032

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, *4*(2), 24.

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique* (Second ed.). Cambridge, MA, USA: MIT Press.

Makeig, S., Gramann, K., Jung, T.-P., Sejnowski, T. J., & Poizner, H. (2009). Linking

brain, mind and behavior. *International Journal of Psychophysiology*, *73*(2), 95–100. doi: http://dx.doi.org/10.1016/j.ijpsycho.2008.11.008

Mecacci, G., & Haselager, P. (2019a). Five criteria for assessing the implications of NTA technology. *AJOB Neuroscience*, *10*(1), 21–23. doi: 10.1080/21507740.2019.1595781

Mecacci, G., & Haselager, P. (2019b). Identifying criteria for the evaluation of the implications of brain reading for mental privacy. *Science and Engineering Ethics*, *25*(2), 443–461. doi: 10.1007/s11948-017-0003-3

Middendorf, M., McMillan, G., Calhoun, G., & Jones, K. S. (2000). Brain-computer interfaces based on the steady-state visual-evoked response. *IEEE Transactions on Rehabilitation Engineering*, *8*(2), 211–214. doi: 10.1109/86.847819

Mladenovic, J., Frey, J., Joffily, M., Maby, E., Lotte, F., & Mattout, J. (2019). Active inference as a unifying, generic and adaptive framework for a P300-based BCI. *Journal of Neural Engineering*. doi: 10.1088/1741-2552/ab5d5c

Mousavi, M., & de Sa, V. R. (2019). Spatio-temporal analysis of error-related brain activity in active and passive brain-computer interfaces. *Brain-Computer Interfaces*, *6*(4). doi: 10.1080/2326263X.2019.1671040

Mousavi, M., Koerner, A. S., Zhang, Q., Noh, E., & de Sa, V. R. (2017). Improving motor imagery BCI with user response to feedback. *Brain-Computer Interfaces*, *4*(1–2), 74–86. doi: 10.1080/2326263X.2017.1303253

Mühl, C., Jeunet, C., & Lotte, F. (2014). EEG-based workload estimation across affective contexts. *Frontiers in Neuroscience*, *8*, 114. doi: 10.3389/fnins.2014.00114

Mullen, T. R., Kothe, C. A. E., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., ... Cauwenberghs, G. (2015). Real-time neuroimaging and cognitive monitoring using wearable dry EEG. *IEEE Transactions on Biomedical Engineering*, *62*(11), 2553–2567. doi: 10.1109/TBME.2015.2481482

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. doi: 10.1016/j.clinph.2007.04.026

Negri, P., Gamberini, L., & Cutini, S. (2014). A review of the research on subliminal techniques for implicit interaction in symbiotic systems. In G. Jacucci, L. Gamberini, J. Freeman, & A. Spagnolli (Eds.), *Symbiotic interaction* (pp. 47–58). Cham, Switzerland: Springer International Publishing. doi: 10.1007/978-3-319-13500-7_4

Parasuraman, R. (2003). Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science*, *4*(1–2), 5–20. doi: 10.1080/14639220210199753

Parmentier, F. B., Elsley, J. V., Andrés, P., & Barceló, F. (2011). Why are auditory novels distracting? contrasting the roles of novelty, violation of expectation and stimulus change. *Cognition*, *119*(3), 374–380. doi: 10.1016/j.cognition.2011.02.001

Perera, C., Zaslavsky, A., Christen, P., & Georgakopoulos, D. (2014). Context aware computing for the internet of things: A survey. *IEEE Communications Surveys Tutorials*, *16*(1), 414–454. doi: 10.1109/SURV.2013.042313.00197

Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology*, *40*(1), 187–195. doi: 10.1016/0301-0511(95)05116-3

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners.

Ramoser, H., Müller-Gerking, J., & Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Transactions on*

Rehabilitation Engineering, *8*(4), 441–446.

Rosenfeld, J. P., Cantwell, B., Nasman, V. T., Wojdac, V., Ivanov, S., & Mazzeri, L. (1988). A modified, event-related potential-based guilty knowledge test. *International Journal of Neuroscience*, *42*(1–2), 157–161. doi: 10.3109/00207458808985770

Schmidt, A. (2000). Implicit human computer interaction through context. *Personal and Ubiquitous Computing*, *4*(2-3), 191–199. doi: 10.1007/BF01324126

Schreuder, M., Höhne, J., Blankertz, B., Haufe, S., Dickhaus, T., & Tangermann, M. (2013). Optimizing event-related potential based brain–computer interfaces: a systematic evaluation of dynamic stopping methods. *Journal of Neural Engineering*, *10*(3), 036025. doi: 10.1088/1741-2560/10/3/036025

Schultheis, H., & Jameson, A. (2004). Assessing cognitive load in adaptive hypermedia systems: Physiological and behavioral methods. In P. M. E. De Bra & W. Nejdl (Eds.), *Adaptive hypermedia and adaptive web-based systems* (pp. 225–234). Berlin Heidelberg: Springer. doi: 10.1007/978-3-540-27780-4_26

Senden, M., Emmerling, T. C., van Hoof, R., Frost, M. A., & Goebel, R. (2019). Reconstructing imagined letters from early visual cortex reveals tight topographic correspondence between visual mental imagery and perception. *Brain Structure and Function*, *224*(3), 1167–1183. doi: 10.1007/s00429-019-01828-6

Settles, B. (2009). *Active learning literature survey* (Computer Sciences Technical Report No. 1648). Madison, WI, USA: University of Wisconsin–Madison.

Sirevaag, E. J., Kramer, A. F., Coles, M. G., & Donchin, E. (1989). Resource reciprocity: An event-related brain potentials analysis. *Acta Psychologica*, *70*(1), 77–97. doi: 10.1016/0001-6918(89)90061-9

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 2951–2959). Red Hook, NY, USA: Curran Associates.

Solovey, E., Schermerhorn, P., Scheutz, M., Sassaroli, A., Fantini, S., & Jacob, R. (2012). Brainput: Enhancing interactive systems with streaming fNIRS brain input. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2193–2202). New York, NY, USA: ACM. doi: 10.1145/2207676.2208372

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* Cambridge, MA, USA: MIT Press.

Treder, M. S., Schmidt, N. M., & Blankertz, B. (2011). Gaze-independent brain-computer interfaces based on covert attention and feature attention. *Journal of Neural Engineering*, *8*(6), 066003. doi: 10.1088/1741-2560/8/6/066003

Vidal, J. J. (1977). Real-time detection of brain events in EEG. *Proceedings of the IEEE*, *65*(5), 633–641.

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: An electric sign of sensori-motor association and expectancy in the human brain. *Nature*, *203*(4943), 380–384. doi: 10.1038/203380a0

Wenzel, M. A., Bogojeski, M., & Blankertz, B. (2017). Real-time inference of word relevance from electroencephalogram and eye gaze. *Journal of Neural Engineering*, *14*(5), 056007. doi: 10.1088/1741-2552/aa7590

Wirth, C., Dockree, P. M., Harty, S., Lacey, E., & Arvaneh, M. (2019). Towards error categorisation in BCI: single-trial EEG classification between different errors. *Journal of Neural Engineering*, *17*(1), 016008. doi: 10.1088/1741-2552/ab53fe

Yeung, N., & Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *The Journal of Neuroscience, 24*(28), 6258–6264.

Yuksel, B. F., Oleson, K. B., Harrison, L., Peck, E. M., Afergan, D., Chang, R., & Jacob, R. J. K. (2016). Learn piano with BACh: An adaptive learning interface that adjusts task difficulty based on brain state. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 5372–5384). New York, NY, USA: ACM. doi: 10.1145/2858036.2858388

Yuste, R., Goering, S., Agüera y Arcas, B., Bi, G., Carmena, J. M., Carter, A., ... Wolpaw, J. (2017). Four ethical priorities for neurotechnologies and AI. *Nature, 551*, 159–163. doi: 10.1038/551159a

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

Zander, T. O., Brönstrup, J., Lorenz, R., & Krol, L. R. (2014). Towards BCI-based Implicit Control in Human-Computer Interaction. In S. H. Fairclough & K. Gilleade (Eds.), *Advances in Physiological Computing* (pp. 67–90). Berlin, Germany: Springer. doi: 10.1007/978-1-4471-6392-3_4

Zander, T. O., & Jatzev, S. (2009). Detecting affective covert user states with passive brain-computer interfaces. In *3rd International Conference on Affective Computing and Intelligent Interaction (ACII) and Workshops* (pp. 1–9). doi: 10.1109/ACII.2009.5349456

Zander, T. O., & Jatzev, S. (2012). Context-aware brain-computer interfaces: exploring the information space of user, technical system and environment. *Journal of Neural Engineering, 9*(1), 016003.

Zander, T. O., & Kothe, C. A. (2011). Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *Journal of Neural Engineering, 8*(2), 025005. doi: 10.1088/1741-2560/8/2/025005

Zander, T. O., Krol, L. R., Birbaumer, N. P., & Gramann, K. (2016). Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity. *Proceedings of the National Academy of Sciences, 113*(52), 14898–14903. doi: 10.1073/pnas.1605155114