Inferring Causality from Noninvasive Brain Stimulation in Cognitive Neuroscience

Til Ole Bergmann¹ and Gesa Hartwigsen²

Abstract

Noninvasive brain stimulation (NIBS) techniques, such as transcranial magnetic stimulation or transcranial direct and alternating current stimulation, are advocated as measures to enable causal inference in cognitive neuroscience experiments. Transcending the limitations of purely correlational neuroimaging measures and experimental sensory stimulation, they allow to experimentally manipulate brain activity and study its consequences for perception, cognition, and eventually, behavior. Although this is true in principle, particular caution is advised when interpreting brain stimulation experiments in a causal manner. Research hypotheses are often oversimplified, disregarding the underlying (implicitly assumed) complex chain of causation, namely, that the stimulation technique has to generate an electric field in the brain tissue, which then evokes or modulates neuronal activity both locally in the target region and in connected remote sites of the network, which in consequence affects the cognitive function of interest and eventually results in a change of the behavioral measure. Importantly, every link in this causal chain of effects can be confounded by several factors that have to be experimentally eliminated or controlled to attribute the observed results to their assumed cause. This is complicated by the fact that many of the mediating and confounding variables are not directly observable and dose-response relationships are often nonlinear. We will walk the reader through the chain of causation for a generic cognitive neuroscience NIBS study, discuss possible confounds, and advise appropriate control conditions. If crucial assumptions are explicitly tested (where possible) and confounds are experimentally well controlled, NIBS can indeed reveal cause–effect relationships in cognitive neuroscience studies.

INTRODUCTION

Noninvasive brain stimulation (NIBS) techniques, such as TMS or transcranial direct and alternating current stimulation (TDCS/TACS), allow to experimentally manipulate neuronal activity in the healthy human brain in a temporally and spatially specific manner, thereby overcoming the merely correlational nature of electrophysiological and neuroimaging techniques (Bergmann, Karabanov, Hartwigsen, Thielscher, & Siebner, 2016). Their ability to bypass sensory input channels and directly affect brain activity makes them unparalleled tools for studying cause–effect relationships between neuronal activity and cognitive function. Shortly after its invention (Barker, Jalinous, & Freeston, 1985), TMS was already demonstrated to be capable of suppressing visual perception (Amassian et al., 1989), and by now, the “disruptive” or “interfering” effects of TMS have a long-standing tradition in cognitive neuroscience, following the so-called virtual lesion approach (Pascual-Leone, Walsh, & Rothwell, 2000; Walsh & Cowey, 2000). Later, TDCS (Nitsche & Paulus, 2000) and TACS (Antal et al., 2008) were discovered as methods for subthreshold modulation of neuronal activity and thus of cognitive function (Vosskuhl, Strüber, & Herrmann, 2018; Yavari, Jamil, Mosayebi Samani, Vidor, & Nitsche, 2018). Although cognitive neuroscience studies using these NIBS techniques often claim to test the “causal relevance” of a specific brain region or neuronal activity pattern for a specific cognitive function or behavior, the underlying cause–effect relationships are rarely made explicit. However, to justify such causal inference, the theoretically assumed chain of causation, leading from the applied stimulation to the observed behavioral change, has to hold for a concrete empirical experiment. Crucially, every single cause–effect link in this causal chain can be interrupted or confounded by several factors, which are best eliminated or controlled experimentally to attribute the observed results to their assumed cause. We will start by briefly outlining the core element of this paper: a simplified five-step chain of causation for cognitive neuroscience NIBS studies and its principal confounders. We will then introduce general experimental approaches using NIBS and discuss the concept of causal inference for the case of experimental NIBS studies in cognitive neuroscience before we walk the reader step-by-step through the five-step chain of causation. Afterward, we will discuss potential confounders in
The Chain of Causation in NIBS Studies

A chain of causation (or causal chain) refers to an uninterrupted concatenation of cause–effect pairs, leading from an initial cause of interest via a number of mediating variables to an eventual effect of interest. Given that most effects have many causes and themselves cause many effects, such a chain represents only one specific path through an entire causal diagram. A causal diagram can be formalized as variables connected with arrows that indicate causation (A → B) instead of mere association (A, B), with the left and right variables representing cause and effect for a particular cause–effect pair (Pearl & Mackenzie, 2018). Importantly, there is no ultimate cause or effect, and the partial chain to be considered depends entirely on the research question at hand. Once the hypothesis has defined cause and effect of interest, any intermediate element within a causal path connecting them is referred to as “mediator.” In contrast, elements that are associated with the cause of interest (i.e., causing it or merely covarying with it but not caused by it) and that influence the effect of interest are considered confounders, because their influence on the effect of interest is mixed with that of the cause of interest, preventing the straightforward attribution of causal influence to the latter and therefore the identification of an unambiguous causal path through the causal diagram.

The somewhat naïve level at which hypotheses are often phrased in NIBS studies is depicted in Figure 1A. The stimulation is expected to affect a single circumscribed brain region, which has an effect on behavior, because the brain region is causally relevant for producing that behavior. In Figure 1B, we provide a more elaborated causal diagram, but for the sake of comprehensibility, it is still simplified and reduced to the key variables (some of them summarizing multiple smaller ones). The red elements connected by red arrows indicate the core chain of causation leading from application of NIBS to an observable behavioral effect. The yellow box and arrows indicate the causal route by which task demands and the current brain state drive local and network activity and thereby the respective cognitive function and behavioral responses. This causal path is not only a source of confounding but is relevant for task performance in the absence of NIBS and the main drive for the cognitive function of interest. In fact, NIBS-induced brain activity per se is insufficient to cause more complex cognitive functions and only modulates ongoing task-related neuronal activity. The black arrows indicate additional causal relationships that result from or eventually affect elements of the main causal chain of interest and may thereby produce confounding via central (blue boxes) and peripheral (green boxes) off-target stimulation. We segregate the core chain of causation from NIBS application to behavior into five cause–effect pairs as described below.

Arrow 1: The application of NIBS produces an electric field (E-field) in the brain tissue, either via transcranial electro-magneto-electric induction (TMS) or the direct administration of weak transcranial currents (transcranial current stimulation [TCS]).

Arrow 2: The E-field then interacts with the neurons’ membrane potential to immediately (online) evoke neuronal firing (TMS) or modulate the threshold for spontaneous firing (TCS) locally in the targeted brain region, activating specific intracortical circuit motifs and associated neuronal signatures (Womelsdorf, Valiante, Sahin, Miller, & Tiesinga, 2014). In the case of offline TMS/TCS protocols, it additionally triggers processes of synaptic plasticity.

Arrow 3: If driving output neurons of the targeted network node to suprathreshold levels, local neuronal activity transsynaptically spreads to other connected brain regions of the targeted network via intercortical axonal projections, activating large-scale and remote circuit motifs as well as remote changes in synaptic strength.

Arrow 4: The immediate (online) or subsequent (offline) effects on local and large-scale circuit motifs can disturb or facilitate the specific task-relevant neuronal computations mediating the cognitive function of interest, which is reflected in either respective changes of the outcome or the completion time of these processes.

Arrow 5: With regard to the motor responses exerted in the context of a specific behavioral task, these altered cognitive processes eventually result in changed error rates or RTs.

Confounding in NIBS Studies

Most cognitive neuroscience studies investigate the neuronal implementation of a cognitive function and therefore...
hypothesize that neuronal activity in a certain brain region is causal to that cognitive function of interest. NIBS is merely considered a means to manipulate the cause "neuronal activity" (via an E-field), whereas behavioral measures are used as an observable proxy to assess the hidden effect "cognitive function." Every single link in this causal path can be confounded by several (known or unknown) factors (yellow, blue, and green boxes in Figure 1B), which are best experimentally eliminated or controlled for to rule out alternative explanations for the observed data and to draw strong conclusions regarding the hypothesized cause–effect relationship. This is complicated by the facts that (i) many of the mediating and confounding variables are not directly observable and have to be approximated by simulation (E-field), neuroimaging (neuronal activity), or modeling (cognitive function) and (ii) few of the relevant cause–effect pairs express linear dose–response relationships. Depending on the expertise and educational background of the researcher, some of these links are typically less thoroughly elaborated than others (e.g., the kind of neuronal activity induced by the E-field or the precise behavioral changes to be expected from changes in cognitive function), crucial assumptions about mediators remain untested (e.g., whether the TMS-induced E-field actually produced the neuronal activity that was aimed for), and potential confounds are uncontrolled (e.g., the sensory input because of peripheral costimulation).

**GENERAL APPROACHES FOR NIBS IN COGNITIVE NEUROSCIENCE**

Although NIBS is often depicted as a means by which we can simply "switch off" or "knock out" entire brain regions or realistically mimic endogenous oscillatory activity, the
Experimental Approaches Using NIBS

Depending on their specific stimulation parameters, both TMS and TCS are able to exert not only immediate (online) effects during stimulation but also subsequent (offline) effects that outlast the stimulation itself for minutes to hours (Figure 2; for a more detailed discussion of NIBS approaches, see Bergmann et al., 2016). Online approaches, assessing the immediate neural response to stimulation, can be used to (i) quantify properties such as cortical excitability or connectivity and their modulation by brain state (including task engagement; Figure 2A), (ii) interfere with ongoing task-related or spontaneous neuronal activity and thereby with cognition (Figure 2B), or (iii) modulate, more gently, the level (“gating”) and timing (“entrainment”) of neuronal activity and thereby cognitive function (Figure 2C). In contrast, offline approaches can be utilized to either (iv) inhibit or (v) facilitate neuronal excitability for an extended period via mechanisms of synaptic plasticity, assessing its subsequent effects on neuronal activity and cognition (Figure 2D and E). Although relying on different neurophysiological mechanisms (which we will discuss in a later section), both online interference and offline inhibition approaches have been referred to as “virtual lesion” and are most frequently used in cognitive neuroscience to test whether and when a cortical region is “causally relevant” for a cognitive function. Online modulation additionally allows to investigate which neuronal patterns, for example, oscillatory frequencies, are mediating or supporting a specific cognitive function. Whereas TMS works for all these approaches, the much weaker TCS cannot be used for quantification (as it does not trigger APs) and hardly for interference. In contrast, it is optimally suited to modulate without disruption and can produce lasting offline effects. We will see that net increases and decreases in cortical excitability (online as well as offline) do not simply translate into respective behavioral improvements and impairments.

CAUSAL INFEERENCE FROM NEUROIMAGING VS. NIBS IN COGNITIVE NEUROSCIENCE

Causal inference refers to the process of inferring cause–effect relationships based on the observed changes of an
effect, after changes of its hypothesized cause. In contrast to inferring a mere association of the two variables by means of conditional probabilities, that is, $P(\text{effect} \mid \text{cause})$, causal inference assigns a direction to their relationship, assuming that active manipulation of the cause (experimentally or counterfactually), with everything else held constant (“ceteris paribus”), produces the effect, but not vice versa. This asymmetric relationship has been formalized via the do operator, where $P(\text{effect} \mid \text{cause}) = P(\text{effect} \mid \text{do(cause)})$, whereas $P(\text{cause} \mid \text{effect}) = P(\text{cause} \mid \text{do(effect)}) = P(\text{cause})$ (Pearl & Mackenzie, 2018). A simple example in the context of NIBS would be that TMS (with sufficient intensity) of the primary motor cortex (M1) hand area causes a contralateral finger movement ~20 msec later or at least increases its likelihood ($P(\text{contraction} \mid \text{do(TMS)})$, but a spontaneous finger movement does not affect the likelihood of TMS to occur ($P(\text{TMS} \mid \text{do(contraction)}) = P(\text{TMS})$). In fact, this asymmetry holds for every single cause–effect pair in the causal chain mediating the effect from TMS pulse to finger movement (i.e., TMS pulse $\rightarrow$ E-field $\rightarrow$ APs in cortical neurons in M1 Layer 2/3 or premotor cortex $\rightarrow$ APs in corticospinal output neurons in M1 Layer 5 $\rightarrow$ APs in spinal motoneurons $\rightarrow$ muscular APs $\rightarrow$ muscle contraction $\rightarrow$ finger movement). For behavioral task performance in cognitive neuroscience studies, the causal chain is typically even more complex, and for the causal diagram in Figure 1B, the numerous single steps have been consolidated into a few categories for the sake of simplification.

It has been argued that causal relationships can provisionally be inferred from observational data alone within the causal inference framework when using the do-calculus to decide which confounding variables should be statistically adjusted for and which should rather not to avoid the introduction of spurious effects (Pearl & Mackenzie, 2018; Pearl, 2010). This approach has gained increasing interest in the functional neuroimaging community as well, where its applicability is fiercely debated (Reid et al., 2019; Mehler & Kording, 2018; Grosse-Wentrup, Janzing, Siegel, & Schölkopf, 2016; Weichwald et al., 2015; Ramsey et al., 2010). In contrast, the classical experiment solves this caveat elegantly also with respect to unknown confounders via the randomized allocation of levels of the independent variable (IV) to observational units across multiple experimental repetitions (i.e., do [cause]), while observing the resulting changes in the dependent variable (DV) as a function of the IV level (i.e., $P(\text{effect} \mid \text{do(cause)})$). Although manipulation of the IV and resulting changes of the DV relate to cause and effect for the case of a randomized controlled trial with a single cause–effect pair, it is more ambiguous for more complex chains of causation with multiple cause–effect pairs lined up (like in Figure 1B), where only the first cause (here, application of stimulation) relates to the experimental manipulation of an IV and only the last effect (here, behavior) is assessed as DV, whereas all intermediate steps are hidden variables that can neither be manipulated nor directly be observed. We will thus use the terms “cause” and “effect” within the causal inference framework, but “IV” and “DV” when taking the experimentalist’s perspective.

**Mapping Correlational Relationships with Noninvasive Neuroimaging**

In neuroimaging studies, the cognitive function of interest is isolated experimentally as the only difference between respective task conditions (levels of the IV), whereas the associated neural activity is assessed (as DV) and contrasted between conditions. By experimental variation of the IV in randomized controlled trials, clever design of task conditions, and careful experimental control of possible confounders, such experiments, unlike mere observational studies, allow to attribute the observed DV change to variations of the IV. Nonetheless, the causal direction between cognitive function and brain activity is not easily derived from such experiments. Commonly, this allocation of IV and DV does not imply that a certain mental state is expected to cause the respective brain state ($P(\text{brain state} \mid \text{do(mental state)})$) but rather the opposite ($P(\text{mental state} \mid \text{do(brain state)})$) or, at least, that the noncausal relation of supervenience is assumed (Dijkstra & de Bruin, 2016). While trying to avoid taking any particular philosophical position with respect to the mind–body problem (for an introduction, see Chambliss, 2018; Nagel, 1993), we will, for the ease of argument, assume here that a cognitive function causally depends on a specific neuronal substrate (i.e., the structure) and temporospatial patterns of neuronal activity it produces ($P(\text{cognitive function} \mid \text{do(brain activity)})$). This should not be misunderstood as a dualist view on mental causation, because “cognitive function” does not refer to a conscious, phenomenal experience or mental state but rather pragmatically to the mechanisms of information processing and the computations that eventually give rise to a certain behavior, in the following generously spanning anything from perception, via higher-order cognitive processes, to motor function. From that perspective, a cognitive neuroimaging experiment thus engages the participant in a task requiring for its completion the recruitment of certain neuronal networks and mechanisms implementing the respective target cognitive function, while measuring brain activity as a function of task condition (for a detailed discussion, see Dijkstra & de Bruin, 2016). However, because not all associated brain activities may be causally contributing to the cognitive function engaged by the task, the measured neuronal activity cannot qualify unambiguously as its cause but merely as its neuronal correlate.

**Mapping Causal Relationships with NIBS**

This ambiguity of causal direction can be resolved when using NIBS to experimentally manipulate brain activity.
independent of task engagement instead of merely observing it via neuroimaging, while measuring behavioral task performance as a proxy for the integrity of a cognitive function (Sack, 2006; for a detailed discussion, see Dijkstra & de Bruin, 2016). Therefore, NIBS-related changes in behavior can principally be interpreted as causal effects of the experimentally induced change in brain activity ($P["cognitive function"] do["brain activity"]$). There are few examples when NIBS alone is sufficient to produce motor behavior or a perceptual phenomenon, such as the well-established induction of muscle responses as quantified by the motor-evoked potentials (MEPs) or phosphene (illusory perceptions of light) after TMS of the primary motor and visual cortex, respectively (Kammer, 1998; Mills, Boniface, & Schubert, 1992). For these cases, the online quantification approach (Figure 2A) can be employed to demonstrate a clear causal relationship between neuronal activity in the respective brain structure (e.g., TMS-induced firing of corticospinal motor neurons in M1) and behavioral outcome (e.g., contraction of a contralateral hand muscle), and even dose–response relationships can be identified (e.g., the sigmoidal function relating increasing TMS intensity to increasing MEP amplitude). However, with NIBS in humans, most other brain targets do not result in overt outputs. Instead, behavioral tasks are required to engage the cognitive function of interest and its neuronal correlate, while using NIBS to manipulate the activity or excitability of a target brain region before (offline; Figure 2D and E) or during (online; Figure 2B and C) task performance to reveal its causal contribution to this cognitive function. This brings us back to the heart of this paper: the chain of causation that is tacitly hypothesized for NIBS studies in cognitive neuroscience and its many possible confounders that complicate the causal interpretation of NIBS results (Figure 1B). In the next sections, we will walk the reader step-by-step through the causal path and discuss for each step under which conditions a cause–effect relationship can be assumed. Afterward, we will describe how these causal links can be confounded by variables that systematically covary with the cause and constitute an alternative cause for its effect. Only if an uninterrupted chain of causation can be established without confounding causes for the individual links, the conclusion can be drawn that “neuronal activity in region X is causing behavior Y.”

FROM NIBS APPLICATION TO E-FIELDS (ARROW 1)

The first cause–effect pair (Arrow 1) is often implicitly assumed without further discussion, namely, that the applied NIBS technique produces an E-field of desired intensity, extent, and direction in the target brain region, without affecting nontarget brain regions and without the target brain region being inadvertently affected by other factors associated with the stimulation. This first crucial step is far from trivial, and simply holding a TMS coil or attaching a TCS electrode over the assumed target region is not sufficient for many reasons.

Identifying the Target

Before we can attempt to stimulate a specific target site of interest, we need to determine its location. Depending on the spatial specificity of the NIBS method, these targets are easily underspecified or overspecified. For TMS, effective current densities are restricted to less than a cubic centimeter (Brasil-Neto, McShane, Fuhr, Hallett, & Cohen, 1992), and targets such as the “posterior parietal cortex” or “dorsolateral pFC” are very unspecific, given that (i) the functional organization of most brain areas is topographically more fine-grained and (ii) only a portion of that anatomical structure will receive effective stimulation. For TCS, in contrast, even entire brain regions can hardly be stimulated in isolation. Irrespective of the spatial specificity of the NIBS technique, the target site can principally be determined based on (i) its function, (ii) its neuroanatomical location, or (iii) even its location relative to the skull alone. A functional TMS localizer can be used to determine a motor or phosphene hot spot based on MEPs and phosphenes, respectively. This approach is highly specific as it allows fine-tuned coil positioning based on the immediate feedback from the output variables, ensuring that the intended neuron population is effectively stimulated. Unfortunately, this method is only available for very few targets (i.e., motor and visual cortex; see Bergmann et al., 2016). The second-best option is a localizer via fMRI, for example, to determine the FEF from a covert spatial attention task (Marshall, O’Shea, Jensen, & Bergmann, 2015) or the extrastriate body area from contrasting body parts versus other objects (Zimmermann, Verhagen, de Lange, & Toni, 2016). While providing no information on coil orientation, the target voxel can be determined based on individual statistical maps. For functions tightly linked to an identifiable anatomical location (e.g., motor hand knob), an individual structural MRI scan alone may be used to identify the target coordinates, yet allowing considerable uncertainty within that area. When ignoring interindividual variability in a structure–function relationship, standard coordinates from the literature can be utilized after transforming them to native space with the help of an individual structural MRI (Duecker et al., 2014). When disregarding individual brain anatomy altogether, the 10–20 EEG electrode system can be used to roughly estimate the location of specific brain regions (e.g., F3 for the left dorsolateral pFC or P4 for the right posterior parietal cortex). Systematic comparisons revealed that, with decreasing individualization across the above-described methods, the number of participants required to observe a significant effect increases dramatically (Sack et al., 2009; Sparing, Buelte, Meister, Pauš, & Fink, 2008). Importantly, all approaches besides the TMS localizer
and the 10–20 system require an MR-informed frameless stereotactic neuronavigation setup to position the TMS coil over the target site, which is considered state-of-the-art to maintain coil position within and across experimental sessions.

### Reaching the Target

Once we know where to stimulate, how can we ensure that the desired E-field is expressed in the target site? For TMS, a high-voltage current pulse (< 300 μs), running through an insulated coil held tangential to the scalp, produces a magnetic field that painlessly penetrates the skull and in turn produces an electric current in the underlying brain tissue. Importantly, the magnetic field is not attenuated by the intermediate bone, but the induced E-field in the brain simply decays exponentially with distance from the TMS coil (Thielscher & Kammer, 2004). This highlights the role of the scalp–cortex distance, which is known to vary across both brain regions and individuals and can partially be accounted for by adjusting stimulation intensity to the actual scalp–cortex distance (Stokes et al., 2005). In addition, the local E-field distribution depends on the anatomical distribution of brain tissues with different conductivities (gray matter, white matter, corticospinal fluid) and the individual gyriﬁcation of the underlying cortex (Opitz, Windhoff, Heidemann, Turner, & Thielscher, 2011; Thielscher, Opitz, & Windhoff, 2011). The E-field induced by TCS, in contrast, has to pass through the bone, which is a major barrier of low conductivity, causing large portions of the stimulation current to be shunted via the scalp and to enter the skull via openings, such as eyeballs, ear canals, or small foramen for the cranial nerves. Within the brain, the E-field distribution depends again on the distribution of brain tissues, but unlike for TMS, the TCS-related E-field extends across a much larger brain volume (depending on the specific electrode type and montage; Opitz, Paulus, Will, Antunes, & Thielscher, 2015; Datta et al., 2009). Importantly, E-field simulations based on anatomically precise individual head models revealed that the location of the maximum E-field varies across brain regions and individuals and is not simply located directly underlying the TMS coil (Weise, Numssen, Thielscher, Hartwigsen, & Knösche, 2020; Gomez-Tames, Hama-Saka, Laakso, Hirata, & Ugawa, 2018) or TCS electrode (Opitz et al., 2015; Saturnino, Antunes, & Thielscher, 2015). To establish that an effective stimulation intensity is achieved at the target coordinate, individualized E-field modeling is advisable for both TCS and TDC, although additional mechanisms have been discussed (Peterchev et al., 2012). For TMS, the E-field dynamics are sufficiently fast and strong to depolarize the neuronal membrane to suprathreshold levels, presumably at the level of axons or axon terminals (Abera, Wang, Grill, & Peterchev, 2020), to a degree that APs emerge and spread along the membrane. These APs then transsynaptically affect connected neurons, causing excitatory and inhibitory postsynaptic potentials, via glutamate and GABA-A/GABA-B receptors, depending on the initially depolarized neuron type. Spatial and temporal integration of postsynaptic potentials then causes excitatory and inhibitory postsynaptic neurons to fire. Although the intracortical circuitry responding to TMS has been studied in great detail using paired-pulse protocols and pharmacological interventions in the primary motor cortex (Di Lazzaro & Ziemann, 2013; Di Lazzaro, Ziemann, & Lemon, 2008) and some insights have been generated by work in rodents or nonhuman primates (Romero, Davare, Armendariz, & Janssen, 2019; Li et al., 2017; Mueller et al., 2014), the specific synaptic responses in most human cortical regions can only be speculated about. In any case, TMS evokes highly synchronized neuronal responses of entire intracortical circuits, not only because the E-field initially depolarizes a large number of different neurons but also because the activation spreads among them. Consequently, there will be both excitation and inhibition within a neuronal population or brain region after TMS, and the net effect on its excitability or information processing capabilities is complex (as will be discussed for Arrow 4).

### Inducing Immediate (Online) Effects in Local Neuronal Activity

Membrane polarization is presumably the main mechanism of action for both TMS and TDC, although additional mechanisms have been discussed (Peterchev et al., 2012). For TMS, the E-field dynamics are sufficiently fast and strong to depolarize the neuronal membrane to suprathreshold levels, presumably at the level of axons or axon terminals (Abera, Wang, Grill, & Peterchev, 2020), to a degree that APs emerge and spread along the membrane. These APs then transsynaptically affect connected neurons, causing excitatory and inhibitory postsynaptic potentials, via glutamate and GABA-A/GABA-B receptors, depending on the initially depolarized neuron type. Spatial and temporal integration of postsynaptic potentials then causes excitatory and inhibitory postsynaptic neurons to fire. Although the intracortical circuitry responding to TMS has been studied in great detail using paired-pulse protocols and pharmacological interventions in the primary motor cortex (Di Lazzaro & Ziemann, 2013; Di Lazzaro, Ziemann, & Lemon, 2008) and some insights have been generated by work in rodents or nonhuman primates (Romero, Davare, Armendariz, & Janssen, 2019; Li et al., 2017; Mueller et al., 2014), the specific synaptic responses in most human cortical regions can only be speculated about. In any case, TMS evokes highly synchronized neuronal responses of entire intracortical circuits, not only because the E-field initially depolarizes a large number of different neurons but also because the activation spreads among them. Consequently, there will be both excitation and inhibition within a neuronal population or brain region after TMS, and the net effect on its excitability or information processing capabilities is complex (as will be discussed for Arrow 4).
For TCS, the E-field is much weaker and assumed to merely shift the membrane potential slightly toward depolarization or hyperpolarization, changing neuronal excitability on a subthreshold scale (Liu et al., 2018; Stagg & Nitsche, 2011), either constantly (TDCS) or rhythmically (TACS). The E-fields induced by standard stimulation intensities (1–2 mA) in humans are much lower than those in mice or monkeys (Aleksiejchuk, Mantell, Shirinpour, & Opitz, 2019), and the effectiveness of TCS in humans is thus highly debated (Filmer, Mattingley, & Dux, 2020; Liu et al., 2018; Vöröslakos et al., 2018), although TCS-induced E-fields as small as 0.2–1 V/m have already proven effective in causing tiny shifts in spontaneous neuronal firing rates (Krause, Vieira, Csorba, Pilly, & Pack, 2019; Liu et al., 2018; Reato, Rahman, Bikson, & Parra, 2010). Although small, the TCS-induced E-field is broad, and the effects may accumulate across large neuron populations. Again, entire circuits will be stimulated both directly by the E-field and indirectly via synaptic connections, but the response will be less synchronized and more strongly dependent on ongoing brain activity compared to TMS. Importantly, the impact of the E-field on different neuronal structures at the cellular level depends on both their shape and their orientation with the brain and is thus highly complex, with the net effect on a given neuron depending on the integration of diverse depolarization and hyperpolarization of its parts (Rahman et al., 2013). Even more so, the net effect on an entire brain region arises from the integration of the individual neurons’ excitability changes, highlighting the impossibility of a simple relationship between TCS polarity and the resulting net excitability change of the target brain region.

Quantifying Excitability and Connectivity

The quantification of motor or visual cortical excitability via TMS-induced MEPs or phosphenes comes with the inherent proof of suprathreshold stimulation of target neurons (Figure 2A). There are many studies that elegantly employ MEP or phosphene measurement to demonstrate the modulation of motor or visual cortical excitability under various task conditions (Lepage, Saint-Amour, & Théoret, 2008; Sparing et al., 2002). Using dual-coil TMS, effective connectivity with those brain regions can be assessed in a task-dependent fashion (Murakami, Restle, & Ziermann, 2012; Davare, Lemon, & Olivier, 2008). However, these studies typically use NIBS in a correlative manner, treating target brain region excitability as a DV, not an IV, and do not probe the causal impact of brain activity on cognition.

Interfering with Spontaneous or Task-related Neuronal Activity

Many classic cognitive neuroscience TMS studies aim to interfere with neuronal activity during task processing (online) in a specific brain region to demonstrate its causal relevance for a cognitive function (Figure 2B). Unfortunately, we lack a precise understanding of most neuronal activity patterns implementing a specific computation and thus do not know exactly what to interfere with. Accordingly, TMS for interference often uses either high-intensity single-pulse TMS when aiming for a good temporal resolution and thus short period of interference (Amassian et al., 1998) or short TMS bursts at a high frequency (mostly 10–20 Hz) covering several hundred milliseconds to ensure sufficiently long disruption of neuronal processing (Capotosto, Babiloni, Romani, & Corbetta, 2012; Taylor, Nobre, & Rushworth, 2007). Although the TMS-evoked neuronal activity in local circuits is complex, it can safely be assumed that TMS (i) excites random neural elements (those optimally located relative to the E-field), including those not activated by the task; (ii) results in subsequent suppression of neuronal activity, also in neurons activated by the task, for ~50–150 msec after initial excitation (Li et al., 2017; Moliadze, Zhao, Eysel, & Funke, 2003), potentially because of GABA-B-receptor-mediated inhibition, paralleling the motor cortical phenomena of the cortical silent period (Stetkarova & Kofler, 2013; Chen, Lozano, & Ashby, 1999) and long-interval intracortical inhibition (McDonnell, Orekhov, & Ziemann, 2006; Valls-Sölé, Pascual-Leone, Wassermann, & Hallett, 1992); and (iii) causes highly synchronized neuronal activity in the target region based on the time-locked excitation–inhibition pattern artificially evoked in a comparably large neuron population (Romero et al., 2019). We will discuss possible implications of these neuronal effects for the neuronal computations mediating cognition for Arrow 4. Although stimulation intensity and frequency are likely key parameters for determining a successful interference protocol, there has been no systematic comparison of stimulation intensities, frequencies, and train durations regarding their principal suitability for interference protocols.

Modulating (“Gating”) Neuronal Excitability

Online TDCS is often supposed to modulate the excitability of a certain brain region during a task with the rationale to facilitate task-relevant neuronal processing (Figure 2C). Animal work has indeed shown a polarity-dependent modulation of spontaneous neuronal spiking (Fröhlich & McCormick, 2010; Bindman, Lippold, & Redlarn, 1964). However, given the complexity of neuronal excitability changes, in humans, anodal and cathodal TDCS can generally not be equated with excitability increase and decrease outside the primary motor cortex (M1). This issue is somewhat resolved for transcranial random noise stimulation, composed of various (particularly high, > 140 Hz) TACS frequencies (Terney, Chaieb, Moliadze, Antal, & Paulus, 2008). Even for a fixed polarity, no simple dose–response curve can be observed for TCS.
(Esmaeilpour et al., 2018), and these nonlinear effects complicate the determination of appropriate stimulation dosages. Irrespective of these challenges, a noteworthy approach is to induce excitability changes during a learning task to gate learning-induced synaptic plasticity, which can result in long-lasting effects, not resulting from stimulation but learning-related plasticity (O’Shea et al., 2017; Snowball et al., 2013; Vollmann et al., 2013). This approach effectively increases the low anatomical precision of the TCS by the task-related activation of highly specific circuits.

**Entraining Neuronal Activity (and Oscillations)**

Rhythmic TMS or TACS at a certain frequency is used to entrain neuronal activity with the aim to synchronize and enhance endogenous brain oscillations (Figure 2C) and test their causal role for cognition (Vosskuhl et al., 2018; Antal & Herrmann, 2016; Herrmann, Rach, Neuling, & Strüber, 2013; Thut, Schyns, & Gross, 2011; Thut, Veniero, et al., 2011). However, the underlying neurophysiological assumptions are often not made explicit. Neuronal oscillations in EEG/magnetoencephalography (MEG) reflect the summed potentials/fields from large synchronized neuron populations with parallelly oriented dendritic trees (Cohen, 2017). Their amplitude increases when the postsynaptic activity of more neurons becomes synchronized, and the entrainment of a neuronal oscillation by rhythmic NIBS typically refers to the synchronization of spontaneously but yet independently oscillating neurons (Thut, Schyns, et al., 2011). However, it is also possible that random neuronal activity is entrained instead of an already ongoing endogenous oscillation (Herring, Esterer, Marshall, Jensen, & Bergmann, 2019). Entrainment may also work differently for TMS and TCS. Whereas rhythmic (suprathreshold) TMS may directly evoke waves of synchronized excitation and inhibition, potentially phase-resetting existing oscillatory activity (Herring, Thut, Marshall, & Bergmann, 2015), TACS merely shifts the membrane potential forth and back, biasing spontaneous neuronal firing. The weaker impact of TACS thus likely requires it to be more well targeted, for example, adjusted to the individual frequency of the target oscillation (Vosskuhl et al., 2018; Thut et al., 2017). Unfortunately, direct proof of neuronal entrainment during rhythmic NIBS is currently difficult to impossible because of the strong NIBS-related artifacts (Gebodh et al., 2019; Rogasch et al., 2017; Noury, Hipp, & Siegel, 2016; Herring et al., 2015; Ilmoniemi & Kčič, 2010). However, there have been a few successful attempts using more indirect measures of neuronal entrainment (Herring et al., 2019; Hanslmayr, Matuschek, & Fellner, 2014; Helfrich, Schneider, et al., 2014; Thut, Veniero, et al., 2011). Note that any lasting increase in oscillatory power after TACS reflects synaptic aftereffects (offline) in the oscillation-generating circuits, not ongoing entrainment per se (Vossen, Gross, & Thut, 2015; Zaehle, Rach, & Herrmann, 2010).

**Inducing Aftereffects (Offline) in Local Neuronal Excitability Based on Synaptic Plasticity**

Repetitive TMS (rTMS) or prolonged TDCS can produce transient changes in neuronal excitability, mediated by synaptic plasticity and outlasting the stimulation protocol itself by minutes to hours (Figure 2D and E). Several NIBS protocols have been developed for M1, producing bidirectional changes in corticospinal excitability as indexed by MEP amplitude, primarily depending on the frequency or pattern of rTMS or the polarity of TDCS (Ziemann et al., 2008). In principle, classic high-frequency (~5 Hz) versus low-frequency (~1 Hz) rTMS results in long-term potentiation (LTP)-like and long-term depression (LTD)-like facilitation and inhibition of corticospinal excitability, respectively (Fitzgerald, Fountain, & Daskalakis, 2006), whereas for theta burst stimulation, the specific timing of TMS trains and pauses determines the direction of effects (Huang, Edwards, Rounis, Bhatia, & Rothwell, 2005). Likewise, TDCS in classic M1-contralateral forehead montage produces lasting increases and decreases in corticospinal excitability, respectively, depending on whether anode or cathode overlays M1 (Nitsche & Paulus, 2000). Stimulation intensity and duration are crucial determinants for offline effects with both TMS (Ziemann et al., 2008) and TCS (Nitsche et al., 2008). However, Bonaiuto and Bestmann (2015) emphasized that the “sliding-scale rationale” (assuming the magnitude of cortical excitability increases to scale with stimulation intensity) is incorrect, as nonlinearity has been clearly demonstrated even for M1 (Batsikadze, Moliadze, Paulus, Kuo, & Nitsche, 2013; Moliadze, Atalay, Antal, & Paulus, 2012). For a detailed discussion of the neurophysiological mechanisms mediating NIBS-induced LTP/LTD-like plasticity, the reader is referred to previous reviews (Hoogendam, Ramakers, & Di Lazzaro, 2010; Ziemann et al., 2008). Importantly, these aftereffects show large intraindividual and interindividual variability (as discussed below), often emerge with a certain delay (Huang et al., 2005), and wash out after an unknown duration, typically 30–60 min (Ziemann et al., 2008). Although the effectiveness of NIBS can immediately be assessed via MEP amplitudes for M1, such a manipulation check for other brain regions requires neuroimaging techniques (see below). Although common practice, we cannot assume every NIBS protocol to easily translate from motor to nonmotor regions, and without a manipulation check, we can only hope for the desired excitability effects to occur in the target region.

**Mapping NIBS-related Neuronal Effects with Neuroimaging**

Both online and offline NIBS effects can be assessed in humans with noninvasive neuroimaging techniques.
Likewise, concurrent (Antal, Polania, Schmidt-Samoa, for subthreshold intensities (Bestmann et al., 2003). Baudewig, Siebner, Rothwell, & Frahm, 2003, 2005), even effects in remote but anatomically connected cortical TMS et al., 2016; Bestmann & Feredoes, 2013). Concurrent for various motor and cognitive functions (Bergmann which revealed strong remote effects of TMS in networks connections from the initially excited neural elements can further corroborate the causal link between NIBS-induced neuronal and behavioral effects in terms of dose–response relationships.

FROM LOCAL TO NETWORK EFFECTS (ARROW 3)

The third cause–effect pair (Arrow 3) refers to the impact of NIBS-induced local neuronal activity on other connected nodes of the target network. This spread of activation may be desired or considered a potential confound, but it is in any case an inherent feature of the brain, not a shortcoming of the method. Network effects always need to be considered when attributing changes in cognitive function to NIBS-induced changes in the targeted brain region.

Remote Effects of NIBS

The most direct evidence of transsynaptic spread is the MEP after TMS of M1, which relies on several synaptic connections from the initially excited neural elements in M1 via corticospinal output neurons and spinal motorneurons to the muscle. Yet, corticocortical spread has been demonstrated by combined NIBS–fMRI studies, which revealed strong remote effects of TMS in networks for various motor and cognitive functions (Bergmann et al., 2016; Bestmann & Feredoes, 2013). Concurrent TMS–fMRI studies, applying TMS to M1, FEFs, or intraparietal sulcus, found strong (dose- and state-dependent) effects in remote but anatomically connected cortical and subcortical areas (Ruff et al., 2008; Bestmann, Baudewig, Siebner, Rothwell, & Frahm, 2003, 2005), even for subthreshold intensities (Bestmann et al., 2003). Likewise, concurrent (Antal, Polania, Schmidt-Samoa, Dechent, & Paulus, 2011) and consecutive (Polania, Paulus, & Nitsche, 2012) TCS–fMRI studies reported widespread BOLD effects (Turi, Paulus, & Antal, 2012). Yet, for TCS, direct unfocal stimulation effects (i.e., direct effects of the widespread E-field on neuronal activity outside the target region) are difficult to disentangle from actual network effects (i.e., spread of local E-field-induced changes in neuronal activity to remote regions via long-range axonal projections and synaptic connections). In addition, concurrent TMS–EEG studies report TMS-evoked potentials spreading within the targeted network (Harque et al., 2016; Massimini et al., 2005), and dual-coil TMS studies typically build on this feature when testing effective connectivity between two brain regions (Silvanto, Lavie, & Walsh, 2005; Ferbert et al., 1992). Neuroimaging can be used to read out both immediate (online) effects as well as subsequent (offline) effects mediated via synaptic plasticity (Bergmann et al., 2016). For aftereffects on remote neuronal activity, the question remains, however, whether they are caused by local synaptic plasticity in the target site, subsequently affecting remote activity via changes in functional connectivity or via synaptic plasticity in the remote site itself induced by spread of activity during the stimulation.

Consequences for Network Activity

In any case, both online and offline effects on remote nodes can be functionally relevant. For instance, in a consecutive TMS–fMRI study, the rTMS-induced increase in the inhibitory influence of the stimulated area on a remote node predicted the individual TMS-induced response delay in a language task (Hartwigsen et al., 2017). Inhibitory stimulation effects are thus not restricted to the stimulated area but can affect large parts of the network, also modulating the functional interaction of its elements. Such remote network effects remain hidden in purely behavioral studies if single-site TMS or TCS is used and are usually ignored when drawing conclusions about the causal relevance of the stimulated area for a given task. Yet, network effects are potential confounders, especially when relying on plasticity-inducing offline protocols that leave the brain time for adaptive plasticity in response to the intervention and rapid short-term reorganization of the network. Note that remote effects can be inhibitory, facilitatory, or both in different parts of the network, and the direction of network effects is difficult to predict a priori. NIBS-induced inhibition of a key target node sometimes decreases task-related activity in larger parts of the network (Hartwigsen et al., 2017; O’Shea, Johansen-Berg, Trief, Göbel, & Rushworth, 2007), which in turn disinhibits and increases activity in other network nodes, thereby compensating for the disruption and preventing behavioral effects. Such compensatory upregulation can occur in contralateral homologous regions (O’Shea et al., 2007), ipsilateral network nodes (Hallam, Whitney, Hymers, Gouws, & Jefferies, 2016), and neighboring regions relevant for other cognitive functions,
including domain-general areas (Hartwigsen et al., 2017). This short-term reorganization in response to focal disruption stresses the strong potential for flexible redistribution of resources and the high degree of degeneracy in the brain (Price & Friston, 2002; Edelman & Gally, 2001). Combining NIBS with neuroimaging provides a means of mapping both local and remote network effects at the systems level and relating these effects to changes in behavior.

**FROM NEURONAL (NETWORK) ACTIVITY TO COGNITIVE EFFECTS (ARROW 4)**

The fourth cause–effect pair (Arrow 4) refers to the transition from neuronal network activity to a cognitive effect. The latter is not directly observable but must be operationalized as a specific task to be assessed via behavioral performance. Importantly, there is no one-to-one mapping of brain activity to cognitive functions, as the same region is likely involved in multiple functions and the same cognitive function relies on the interaction of multiple regions. Moreover, NIBS protocols can influence the interaction between network activity and cognitive function but rarely produce a direct behavioral output. The main question of this section is thus whether the desired modulation of local and network activity affects the target cognitive function of interest or is assumed to mediate. Both online and offline NIBS approaches can either facilitate or inhibit a cognitive function (Figure 1B), and although a causal discovery per se (e.g., cortical area X but not Y is causally relevant for a cognitive process A but not B) can be made independently from the estimation of direction, size, or specific function of the cause–effect relationship, the latter is crucial for understanding the neuronal mechanism underlying a cognitive function and for developing theory-based applications. It can thus be considered a key challenge to predict a priori the direction and size of the induced effects for a given NIBS protocol, cognitive function, and experimental setting.

**Impairing vs. Improving Cognitive Functions with NIBS**

Many NIBS studies in cognitive neuroscience rely on the “virtual lesion” approach to map causal relationships between neuronal activity in a given brain region and a cognitive function of interest, assuming that disturbing or inhibiting task-related neuronal activity by an online or offline NIBS protocol will result in impairment of the investigated cognitive function. The online interference approach (Figure 2B) aims at transiently disturbing a cognitive function with TMS during task execution, whereas the offline inhibition approach (Figure 2E) relies on a decrease in cortical excitability during the task, mediated by the preceding weakening of synapses after an inhibitory offline NIBS protocol. In both cases, decreases in task performance are expected. Yet, the term “virtual lesion” is misleading, because TMS does not simply switch off a brain region, and offline and online approaches rely on different neuronal mechanisms (cf. Arrow 2). Moreover, “virtual lesions” can only explain performance impairments, and improvements in response to “inhibitory” protocols are often referred to as “paradoxical facilitation” (Walsh & Cowey, 2000; Kapur, 1996). The alternative rationale for demonstrating causal relevance of a brain region or neuronal activity pattern for a given cognitive function is to facilitate it during task execution and show positive consequences for performance. Again, this is typically tried either with online modulation (Figure 2C), entraining task-relevant oscillations via rhythmic TMS or TACS or increasing immediate cortical excitability via TDCS during a task, or by offline facilitation (Figure 2D). The latter is supposed to induce a lasting increase of spontaneous neuronal activity during a subsequent task via the strengthening of synapses with facilitatory offline NIBS. Because it is generally easier to disturb than to improve an insufficiently understood process, the “facilitatory” approach is used less frequently. NIBS-induced facilitation of behavior is nonetheless tempting, as it opens interesting avenues for therapeutic applications or neuroenhancement. Importantly, the choice of NIBS protocols for a specific study is often built on oversimplified assumptions, which partially explains the many null findings and controversial results. Below, we will discuss the possible mechanisms of action translating neuronal into cognitive effects and highlight some of the modulating factors.

**Impairing a Cognitive Function by Online TMS**

The most effective approach for this aim may be online interference via TMS (Figure 2B), which presumably builds on three neuronal effects (cf. Arrow 2). First, the initial excitation of random neural elements causes neuronal noise in the stimulated circuits (Ruzzoli, Marzi, & Miniussi, 2010; Siebner, Hartwigsen, Kassuba, & Rothwell, 2009). Noise pervades all levels of information processing in the nervous system, from receptor signal transduction to behavioral responses (Faisal, Selen, & Wolpert, 2008). The artificial induction of noise may impair or delay task-relevant neuronal computations because neuronal activity needs to be sampled longer to discriminate signal and noise. Second, the initial excitation is inevitably followed by GABAergic feedback inhibition, suppressing neuronal activity for \(\sim\)50–150 msec after TMS (Inghilleri, Berardelli, Cruccu, & Manfredi, 1993; Haug, Schönle, Knobloch, & Köhne, 1992), interrupting and delaying neuronal processing or even causing signal loss during crucial processing steps. This effect may come closest to the “virtual lesion” idea of silencing neuronal activity. Third, the evoked excitation–inhibition sequence artificially synchronizes larger neuron populations, thereby lowering the number of
possible neuronal activity patterns in the network. This loss of entropy (Shannon & Weaver, 1949) in local neuronal activity reduces the information representation capacity of the synchronized network (Hanslmayr, Staudigl, & Fellner, 2012; Schneidman et al., 2011; Tononi, 2008), leading to a degradation of task-relevant information and a disruption of neuronal computations. This may result in prolonged processing time (because of the need for compensatory iterations or the recruitment of additional processing resources) or even an incorrect outcome of the computation. Importantly, online disruption does not leave the targeted network time for functional reorganization (see below), potentially leading to stronger stimulation effects and simplifying the conclusions in comparison to offline approaches.

**Improving a Cognitive Function via Entrainment by Online TACS or TMS**

To actively improve task-related function, specific assumptions are needed regarding the neuronal mechanisms of action. Comparably simple targets are neuronal oscillations, supposedly underlying a variety of cognitive functions (Buzsáki & Draguhn, 2004). To entrain a neuronal oscillation by means of phasic online modulation via rhythmic TMS or TACS (Figure 2C), not only the brain region or network but also the oscillatory frequency needs to be made explicit. When able to transcranially entrain (and augment) a neuronal oscillation and thereby improve the cognitive function, this provides proof of its causal relevance (Vosskuhl et al., 2018). However, whether an increase in synchronization is beneficial depends on the very mechanism by which the oscillation mediates the relevant neuronal computations. Beyond mere excitability fluctuations (Bergmann, Lieb, Zrenner, & Ziemann, 2019; Bergmann et al., 2012; Schroeter & Lakatos, 2009), oscillations supposedly enable more complex processes, such as interarea (Fries, 2015) and cross-frequency communication via phase–phase or phase–amplitude coupling (Jensen & Colgin, 2007), phase coding (Jensen, Gips, Bergmann, & Bonnefond, 2014; Lisman & Jensen, 2013), and potentially phase-dependent plasticity (Bergmann & Born, 2018). These complex processes are more difficult to optimize, although (multifocal) TACS has successfully been used to produce behaviorally relevant interarea synchronization (Reinhart & Nguyen, 2019) and working memory enhancement (Aleksieichuk, Turi, Amador de Lara, Antal, & Paulus, 2016). However, an increase in synchronization, which may unintentionally also recruit task-irrelevant neurons, or an entrainment-induced phase shift of the endogenous oscillation may also be detrimental to task-relevant neuronal computations. Interestingly, the same TMS protocols (e.g., four to five pulse bursts of 10- or 20-Hz TMS) are often used for both behavioral interference (Hartwigsen, Price, et al., 2010) and neuronal entrainment, for example, of alpha or beta oscillatory activity (Romei et al., 2019), and whether a TMS burst impairs or improves a cognitive function may thus depend on whether or not rhythmically synchronized brain activity in the target network is beneficial for the task.

**Bidirectionally Modulating (Gating) a Cognitive Function with Online or Offline NIBS**

In contrast to disrupting or actively driving the neuronal processes mediating a cognitive function, tonic online modulation via TDCS (Figure 2C) is assumedly able to bidirectionally modulate (decrease or increase) task-related neuronal activity, depending on stimulation polarity (among other factors). As discussed for Arrow 2, there is no straightforward relationship between TDCS polarity, intensity, and net excitability changes. There is also no simple mapping from cortical excitability to cognitive performance for any given NIBS protocol. It is well conceivable that a net excitability increase in the circuits mediating task-relevant computations augments task-related signal and thus boosts signal-to-noise ratio (SNR). Alternatively, increased excitability may also augment spontaneous but task-irrelevant activity, thereby increasing noise and lowering the SNR. The same ambiguity applies also for excitability decreases. Importantly, although additional noise may degrade task-relevant SNR for a well-tuned neuronal representation (e.g., when distinguishing similar stimuli in a discrimination task) and thus impair task performance, noise may be beneficial for other tasks (e.g., by lifting a weak perceptual stimulus above threshold in a detection paradigm). The actual effects of online NIBS depend on the complex interaction of spontaneous and task-induced brain states, specific task demands, participant-specific characteristics, and stimulation parameters (Fertonani & Miniussi, 2017). Similar considerations also apply to the rationale of inhibition or facilitation via offline TMS or TCS protocols (Figure 2D and E), although the bidirectional modulation of cortical excitability after offline NIBS is based on different neuronal mechanisms (cf. Arrow 2). The LTP-like strengthening or LTD-like weakening of synapses in the target network results only indirectly in subsequent increases or decreases in neuronal excitability and respective changes in spontaneous and task-related neuronal activity. It should also be noted that evidence for a bidirectionality of offline effects is based almost entirely on the respective modulation of MEP amplitudes after M1 stimulation and may not easily generalize to other montages and cortical areas (Parkin, Bhandari, Glen, & Walsh, 2019). Importantly, as for the effects of online modulation, excitability changes are not homogenously distributed within the targeted brain circuits, because only a random selection of functionally heterogeneous synapses is affected. Despite a possible net facilitation or inhibition of excitability, random changes in either direction are most likely to produce noise in the neuronal activity patterns generated by these circuits, with the above-
discussed positive or negative consequences for task performance. These considerations are highly relevant for training studies and therapeutic applications of NIBS, which often assume an offline NIBS-induced facilitation to improve behavior in a subsequently performed task (Luber & Lisanby, 2014).

The Paradox of Paradoxical Facilitation

“Paradoxical facilitation” usually refers to an unexpected positive effect of an “interfering” or “inhibitory” NIBS protocol on a cognitive function. As emphasized above, noise is a central concept to explain the cognitive effects of NIBS protocols (Ruzzoli et al., 2010; Siebner, Hartwigsen, et al., 2009). Taking stochastic resonance into account, adding noise to a nonlinear system like the human brain may produce opposite effects. Whereas an appropriate amount of noise can add to the weak neuronal signal of a subthreshold stimulus, elevate it above threshold, and result in behavioral facilitation (Schwarzkopf, Silvanto, & Rees, 2011; Miniussi, Ruzzoli, & Walsh, 2010), exceeding noise levels may rather mask the task-relevant neuronal signal. Importantly, the NIBS-induced activity or neural noise is not totally random (Ruzzoli et al., 2010) and also not independent of the task-induced neural activity or brain state. Thus, depending on the activated neuron population, the induced activity can even be considered both as noise and as part of the signal (Miniussi et al., 2010). If the induced neuronal noise is synchronized with the ongoing relevant activity (Ermentrout, Galán, & Urban, 2008), it may augment the signal (Miniussi, Harris, & Ruzzoli, 2013). In other words, behavioral facilitation may result from an optimum level of noise in the system. Although originally described for online TMS studies (Abrahamyan, Clifford, Arabzadeh, & Harris, 2011), these principles also hold for offline TMS and TCS protocols (cf. Fertonani & Miniussi, 2017). Indeed, offline NIBS applied before task processing may transiently prime activity in the stimulated area to a level that facilitates subsequent task performance, although homeostatic metaplasmativity may lead to opposite effects (see below). Besides positive consequences of noise, paradoxical facilitation may also arise from NIBS-induced inhibition of task-irrelevant areas that compete for resources (“addition-by-subtraction” [Luber & Lisanby, 2014]) or the disruption of distracting stimulus elements, facilitating task-relevant processing (Walsh, Ellison, Battelli, & Cowey, 1998). Finally, stimulation-induced disinhibition of distant connected areas may facilitate cognitive processing (Sandrini, Umiltà, & Rusconi, 2011).

Unfortunately, given the complex interaction between task, brain state, stimulation protocol, and intensity, in most cognitive neuroscience experiments, the exact circumstances under which a given NIBS protocol results, on average, in behavioral impairment or facilitation remain unknown, and the above explanations are mainly used in a post hoc fashion. It is even conceivable that, in some cases, for example, as a consequence of increased neuronal noise, the true direction of the NIBS effect on cognitive performance varies across participants or even within participants across trials. In such a case, the mere increase in variance (beyond measurement noise) after NIBS may be considered evidence for a cause–effect relationship, even when lacking a clear direction. However, such a relationship would be less informative regarding the neuronal mechanisms underlying the cognitive function and more difficult to exploit for therapeutic applications.

Various NIBS-induced neuronal effects can affect task-relevant neuronal computations, although little is known about the factors making a specific neuronal process susceptible to or robust against this influence. NIBS effects are often small, and the brain is capable to compensate for weak disturbances, likely contributing to the null effects observed in many NIBS studies.

Multi-site Approaches to Study Network Interactions

Because all cognitive functions rely on distributed processes organized in large-scale neural networks, there is increasing interest in disrupting several network nodes for a given function in a simultaneous or subsequent fashion to study stimulation-induced network effects on cognitive functions. Multifocal TMS can provide insights into functional network interactions and elucidate their compensatory potential. Functional interactions can be studied either online by simultaneously targeting more than one area (“multi-site” approach) or by combining offline and online TMS over different regions (“condition-and-perturb” approach). Multi-site TMS approaches are particularly suited to map the immediate consequences of the disruption of several brain regions because the acute TMS-induced interference during task performance leaves the system no time to develop adaptive plasticity. This allows to test whether the interference effect over one area may be increased by the simultaneous disruption of other key regions, ipsilateral (Ellison & Cowey, 2009) or contralateral (Hartwigsen, Baumgaertner, et al., 2010) to the stimulation site. In a complementary fashion, the condition-and-perturb approach can be used to study rapid network redistribution and compensation (Hartwigsen, 2018). It combines the plastic aftereffects of offline modulation with the immediate perturbation effects of online interference, following the rationale that offline conditioning of one area may sensitize another network node to the disruptive effect of online interference (e.g., Hartwigsen et al., 2012). In some cases, offline conditioning of a single target area does not affect task-related behavior, whereas additional online disruption of a second area effectively impairs task performance, unmasking the disruptive effect of the offline protocol (Hartwigsen et al., 2015; Sack et al., 2009).
likely explanation is that both areas contribute to the function of interest and that offline conditioning of one area can be compensated by a stronger contribution of the other node, changing their functional weights within the network. The additional online perturbation increases the overall “lesion load” and thus results in cognitive disruption. When combined with neuroimaging, condition-and-perturb approaches can also be used to study rapid reorganization at the network level. O’Shea et al. (2007) demonstrated that offline TMS over the left premotor cortex decreased task-related activity in the stimulated area during an action selection task and induced compensatory upregulation in other areas of the motor network, including the contralateral homologous region. Targeting the “reorganized” homologous premotor cortex with subsequent online TMS impaired task performance, demonstrating the functional relevance of the observed compensatory upregulation. This shows how the combination of neuroimaging and multifocal TMS can provide insight into the compensatory dynamics of task-specific neural networks.

Cognitive Models and Computational Neurostimulation

To map stimulation-induced changes on a cognitive function of interest, a valid cognitive model is mandatory that can be translated into a task. For instance, dual-route models have been used to explain stimulus–response compatibility effects in conflict tasks (Ridderinkhof, 2002a). Such models assume parallel routes for decision-making that can be dissociated behaviorally (i.e., a direct, stimulus-based activation route and a controlled, deliberate response activation route) and converge at the level of response activation processes. The dynamics of these processes can be captured with distributional analyses that map interference effects during decision-making (see next section). Thereby, cognitive models help to decompose abstract and complex constructs into several subcomponents that can be operationalized by specific tasks. NIBS can probe the dynamics between these subcomponents and the functional relevance of different brain regions for these processes. More recently, cognitive models have been complemented by computational neurostimulation approaches, which simulate emergent network dynamics and compare them with real data (Bonaiuto & Bestmann, 2015). Computational models may, for example, more accurately reflect choice dynamics by modeling influences of previous trials to capture choice repetition biases (Bonaiuto, de Berker, & Bestmann, 2016; Hämerer, Bonaiuto, Klein-Flügge, Bikson, & Bestmann, 2016). As humans tend to repeat recent choices in real-life situations, modeling the choice history provides valid estimations of decision-making outside the laboratory (Bonaiuto et al., 2016). Such models generate predictions about the effects of NIBS-induced modulations of network dynamics on behavioral measures. They rely on the assumption that perturbation of the dynamics of a biophysical network model via membrane depolarization affects cognitive function (e.g., value-based decision-making) in a predictable way, by modulating the susceptibility of network dynamics to background noise. Thereby, they allow for predicting large-scale network effects of neurostimulation that can be experimentally validated (Bonaiuto et al., 2016; Hämerer et al., 2016).

FROM COGNITIVE FUNCTION TO BEHAVIORAL RESPONSE (ARROW 5)

The fifth cause–effect pair (Arrow 5) refers to the impact of a cognitive function (or its modulation) on the performance in a specific behavioral task. The cognitive function of interest needs to be operationalized by a specific task to measure how it is affected by a given NIBS protocol, and this task has to be sufficiently difficult to be sensitive enough for the (usually very small) cognitive changes induced by NIBS. Usually, the interaction of several cognitive functions is necessary for task completion, and a control task is needed that differs selectively in the target cognitive component to contrast out the influence of other cognitive processes and establish task specificity.

The Benefit of Behavioral Modeling Approaches

To bridge the gap between cognitive model and behavioral outcome measures, distributional analyses can provide insight into different response strategies (Ridderinkhof, 2002a). One advantage relative to composite measures like mean response speed or task accuracy is that they take the whole response distribution into account (i.e., both correct and incorrect responses) and are more sensitive to experimental dynamics and individual differences in response strategies (van den Wildenberg et al., 2010). They help to overcome the poor statistical sensitivity of composite measures (Voss, Nagler, & Lerche, 2013). Distributional analyses map response strategies, such as the speed–accuracy trade-off, by disentangling whether NIBS-induced interference increases errors for fast responses, indicating a potential emphasis on speed, or rather decreases the overall uncertainty in task processing, resulting in increased errors for slower responses. Such analyses further help to distinguish subprocesses of cognitive theories. For instance, within the framework of dual-process models, they have been used to dissociate the role of direct response activation (based on a target stimulus) and selective suppression of activation based on precues in conflict tasks (Ridderinkhof, 2002b). These processing dynamics are usually lost when relying on composite scores. Distributional analyses have demonstrated TDCS-induced modulations of different error types underlying impulsive responses (Spieser, van den Wildenberg,
hasbroucq, riddervik, & burle, 2015), as well as TMS-induced changes in response strategies during action reprogramming (hartwigen & siebner, 2015; Hartwigen et al., 2012) and conflict paradigms (van campen, kunert, van den wildenberg, & riddervik, 2018). Other behavioral modeling approaches rely on sequential sampling models such as the drift diffusion model (ratcliff, smith, brown, & mckoon, 2016; ratcliff, 1978), which assumes reflect the underlying processes contributing to a particular response distribution and also capture decision biases. Such models are particularly sensitive toward slight adaptations of response strategies that may be overlooked when relying on composite measures (ratcliff & mckoon, 2008; voss & voss, 2007), especially in NIBS studies suffering from relatively small effect sizes (see also hartwigen et al., 2015). Aside from binary choice tasks, sequential sampling models have been adapted for more complex multiple decision models (kohl, spieser, forster, bestmann, & yarrow, 2019) that might better match real-life decisions.

**Composite Measures to Quantify NIBS Effects**

Yet, most NIBS studies rely on composite measures derived from the individual mean response speed or accuracy, which are usually analyzed with ANOVAs or t tests. However, NIBS studies can benefit from mixed models, allowing to model nonlinear individual characteristics and providing more flexibility when handling missing data (kaarre et al., 2018; Payne & Tainturier, 2018). Such approaches are especially useful for longitudinal NIBS designs where missing data for single time points might otherwise lead to participant exclusion. Sometimes, motor responses can also be assessed by electrophysiological means, for example, for MEPs recorded from orofacial muscles during stimulation of motor and premotor areas in speech production tasks (Möttönen, van de Ven, & Watkins, 2014; murakami, restle, & ziemann, 2011). For other tasks, psychometric functions (zazio, Bortoletto, Ruzzoli, Miniussi, & Veniero, 2019; Cattaneo et al., 2011) or response biases (riddle, hwang, cellier, Dhanani, & D’esposito, 2019; Smalle, rogers, & Möttönen, 2015) may be the measure of choice. No matter which behavioral measure is used, NIBS protocols are likely to first affect task efficiency, leading to increased (or decreased) response latencies, because response speed is a more sensitive performance metric than task accuracy (Bonaiuto et al., 2016). However, sometimes, task accuracy is affected without any influences on task efficiency (e.g., ward et al., 2010; Amassian et al., 1998). As noted above, speed and accuracy reflect different cognitive strategies that can be disentangled with cognitive models. With respect to the potential mechanisms related to TMS-induced interference effects on either of the two processes, the severity of the interference effect likely depends on the perceptual threshold. For instance, strong interference may suppress a visual stimulus below the perception threshold (Amassian et al., 1998), resulting in decreased task accuracy that cannot be compensated by increased response speed, and weak visual stimuli closer to the perception threshold might be affected first. In contrast, slight modulations of task-related activity may selectively delay response speed, with increases in latencies preventing effects on task accuracy. The timing of the pulses and the stimulation frequency likely play a crucial role for the outcome of a stimulation protocol. For instance, the absence of any impairments in task accuracy in the presence of strong delays in response speed during a visual discrimination task was explained with the employed 10-Hz rTMS protocol, arguing that TMS might disrupt processing for a brief period within each 100-msec interpulse interval but never completely interferes with the relevant information for discrimination (Ellison & Cowey, 2009). Rather, 10-Hz rTMS may merely delay processing by the summed periods of disruption. A similar reasoning likely explains why a 500-msec stimulation period of 10-Hz rTMS does not result in a 500-msec increase in response speed (Walsh & Cowey, 2000). Yet, to the best of our knowledge, no study has systematically varied the stimulation frequency to investigate whether higher frequencies might affect both discrimination speed and accuracy. Most studies assume that TMS might affect both task speed and accuracy, but it remains unclear whether the modulation of either parameter relies on different neural mechanisms. Cognitive and computational models may help to specify the expected effects on both outcome parameters a priori, as the conceptualization of different subprocesses of a cognitive function helps to dissociate the expected outcomes.

**CONFOUNDING FACTORS CHALLENGING THE ASSUMED CHAIN OF CAUSATION**

As outlined in figure 1B, several factors may confound the hypothesized chain of causation (red boxes), by affecting both sides of the investigated core cause–effect pair, that is, the structure–function (or brain–behavior) relationship, namely, the targeted neuronal activity (supposedly caused by NIBS) and the cognitive function of interest (supposedly causing behavioral task performance). Some factors are associated with the application of NIBS, such as the unintended costimulation of nontarget brain regions, either directly (blue boxes) or via peripheral sensory pathways (green boxes). Other factors (yellow box) do not originate from NIBS application but rather from the experimental setup (e.g., task demands), the participants' predisposition (e.g., current brain state, cognitive abilities, beliefs, and expectations), or an interaction of both (e.g., learning effects). All these factors either directly influence brain activity and cognitive function or via a modulation of their response to NIBS. Showing considerable intrapersonal and/or interindividual variability, unsystematic variance in these
factors can compromise the overall effectiveness of a NIBS protocol to modulate the target cognitive function (producing false negatives), whereas systematic variance across experimental conditions can introduce systematic confounding (false positives) for both within- and between-participants designs. To prevent confounding, these factors either need to be eliminated, kept constant across experimental conditions, or explicitly included as experimental control condition (see next section).

Costimulation of Nontarget Regions and Networks

The application of TMS and TCS can have side effects creating relevant confounding of Arrow 2, namely, an effective E-field in nontarget brain regions and the costimulation of peripheral neuronal structures. If the stimulation intensity is sufficiently large, the TMS-induced E-field will reach effective levels also in adjacent nontarget locations (blue boxes), especially in more superficially located ones. It is thus unlikely to exclusively stimulate a coordinate deep in the sulcus, limiting focal stimulation to the gyral crowns (Siebner, 2020; Thielscher et al., 2011). For TCS, the widespread E-field results in even more extended off-target stimulation at similar or higher intensities compared to the target. Although the E-field is more focal for multielectrode montages with small electrodes as compared to classical two-electrode montages with large electrodes (Datta et al., 2009), the effective E-field will not be confined by the borders of the targeted brain area. In addition, both TMS and TCS also induce effective E-fields in the cranial periphery (green boxes), with magnitudes in the skin being inevitably larger than those in the brain (Asamoah, Khatoun, & McLaughlin, 2019), exciting efferent fibers of the facial nerve innervating the facial muscles (Chen, Chauvette, Skorheim, Timofeev, & Bazhenov, 2012), or afferent fibers of the trigeminal nerve innervating the scalp, face, and meninges (Siebner, Auer, Roeck, & Conrad, 1999; Schmid, Moller, & Schmid, 1995). In particular, the TCS-induced E-field, shunted via highly conductive skin tissue, also extends to peripheral neuronal structures in the retina (Lorenz et al., 2019; Schutter, 2016) and the vestibular system (Kwan, Forbes, Mitchell, Blouin, & Cullen, 2019). In addition to the E-field itself, physical side effects of the stimulation, such as the “click” sound and mechanical vibration generated by the discharging TMS coil, also affect mechanoreceptors in the skin and reach the inner ear via both airborne sound waves and bone conduction. Activation of peripheral sensory structures then causes unintended activation of primary and secondary sensory brain regions and, eventually, also higher-order areas. Thus, both pathways (blue and green) can eventually activate remote regions and affect nontarget cognitive functions, for example, via the modulation of attentional orienting or distraction of working memory content (Duecker & Sack, 2015). Therefore, both local and remote network effects can confound the effect of NIBS-induced target activity on the target cognitive function. For example, when targeting higher cortical association areas like the inferior parietal cortex, which integrates information from several modalities (Seghier, 2013), its NIBS-related activation could result from transcranial or sensory stimulation. Complex domain-specific cognitive functions usually engage several domain-general processes like attention, working memory, or executive functions, and sensory disruption of these processes (or other cognitive functions of no interest) may severely affect task processing and modulate response speed or accuracy.

Participants’ Beliefs and Expectations

Participants in NIBS studies have certain expectations and beliefs about the impact of neurostimulation and potential side effects. These expectations and beliefs can influence their performance in a way that may be congruent or incongruent with the experiment’s underlying hypothesis. Participants may also develop expectations and actual knowledge with regard to the temporal structure of the experiment and the current stimulation condition (effective or sham), such that they may anticipate and prepare for stimulation trials in online NIBS studies or behave differently after an offline NIBS protocol.

Task Demands, Learning Effects, and Cognitive Abilities

Many cognitive functions vary considerably over time (e.g., as a consequence of learning or fatigue) and across participants. Differences in baseline performance levels because of different cognitive abilities (low vs. high performers) contribute to the large interindividual variability observed in NIBS studies of cognition. In visual priming studies, single-pulse TMS facilitated behavioral responses in low performers but delayed response speed in high performers, and these effects interacted with task difficulty, indicating a complex interaction between stimulation, brain state, and task-induced state (e.g., Silvanto, Bona, Marelli, & Cattaneo, 2018; Schwarzkopf et al., 2011). Such interactions complicate the conclusions drawn from a given NIBS protocol. Moreover, learning or generalization effects influence performance, especially if the same task is measured repeatedly under different stimulation conditions, as usually done for within-participant designs. Potential influences of these effects need to be controlled by counterbalancing the order of tasks and stimulation conditions across participants. In some learning paradigms, mirrored or alternative sequences or stimulus lists can be used for a second session. In case of implicit tasks or learning paradigms that cannot be repeated, between-participant designs need to be employed. Other unspecific effects such as time of day and hormonal cycle may also influence cognitive function. Some of these factors can (and should)
be kept constant across experimental conditions, for instance, by performing repeated measures in the same participant at the same time of day.

The Current Brain State

The magnitude and direction of NIBS effects may also vary because of differences in the current brain state at the time of stimulation, both within and between participants (Miniussi et al., 2013). The concept of “state dependency” has been first introduced in the visual system (Silvanto, Muggleton, & Walsh, 2008; Silvanto, Muggleton, Cowey, & Walsh, 2007), and state-dependent effects have been described across a variety of (cognitive) domains (Silvanto & Cattaneo, 2017). An impressive example for state dependency is the quantification of corticospinal excitability via the MEP acquired at rest and under precontraction. When compared to a relaxed target muscle, slight precontraction leads to a considerable increase in the MEP size (Siebner, Hartwigsen, et al., 2009). State dependency has also been demonstrated in remote network nodes by simultaneous TMS–fMRI when comparing the effects of different TMS intensities on neural activity at rest and during a grip task (e.g., Bestmann et al., 2008). The key assumption is that the brain state affects the distribution of excitability in the stimulated population of neurons, which in turn affects their responsiveness to stimulation. A modulation of MEP amplitude has also been shown as a function of the current amplitude and phase of neuronal oscillations during sleep (Bergmann et al., 2012) and wakefulness (Bergmann et al., 2019; Thies, Zrenner, Zieman, & Bergmann, 2018). Such oscillatory modulations of cortical excitability led to the idea of brain-state-dependent brain stimulation, allowing to confine stimulation to a certain target state (Bergmann, 2018). Brain state dynamics may also change the NIBS interference effect, such that intensities, which normally impair perception, suddenly have a facilitatory effect (Silvanto et al., 2018; Silvanto & Cattaneo, 2017). For instance, when applied immediately before or early during a task, TMS resulted in a priming effect by increasing activity in the target area to a level optimal for task performance (e.g., Klaus & Hartwigsen, 2019; Töpper, Mottaghy, Brügmann, Noth, & Huber, 1998). In contrast, TMS impaired performance when given during the same task (Wassermann et al., 1999; Flitman et al., 1998). The current brain state may also interact with the polarity in TDCS studies, which is particularly crucial for learning studies that engage different training phases (Dockery, Hueckel-Weng, Birbaumer, & Plewnia, 2009). Although initially introduced for online TMS, state dependency has also been suggested to influence offline NIBS protocols (Nguyen, Deng, & Reinhart, 2018; Miniussi et al., 2013). The facilitatory aftereffect of 10-Hz TACS on subsequent EEG alpha band power, for example, is only evident when TACS is applied with eyes open but not eyes closed and thus during different alpha oscillation amplitudes (Neuling, Rach, & Herrmann, 2013). In addition, classical offline NIBS protocols show large intraindividual and interindividual variability in their aftereffects, depending on several factors that vary across sessions within an individual (e.g., the current brain state, history of synaptic activity, hormonal levels, circadian rhythms) or across individuals (e.g., sex, age, individual depth and orientation of target region, genetics; Ridding & Ziemann, 2010), sometimes leading to null results (Beaulieu, Flamand, Massé-Alarie, & Schneider, 2017; Heidegger et al., 2017). The individual variability in response to offline TMS protocols in the motor system may be influenced by the specific interneuron networks recruited, and corticospinal excitability measures may help to dissociate responders and nonresponders (Hamada, Murase, Hasan, Balaratnam, & Rothwell, 2013). However, it is unclear how this translates to cognitive functions, because excitability of M1 is not related to the responsiveness of other areas, and interindividual differences in cognitive abilities and functional organization further contribute to large interindividual variability in response to NIBS protocols when targeting cognitive functions.

Metaplasticity

Offline stimulation protocols may further be affected by metaplasticity, such that synaptic plasticity itself may vary depending on the history of a neuron’s postsynaptic activity (Abraham & Bear, 1996). Metaplasticity contributes to network function and behavior and may be homeostatic or nonhomeostatic. Homeostatic metaplasticity has been demonstrated in human M1 for several combinations of plasticity-inducing NIBS protocols (Karabanov et al., 2015; Müller-Dahlhaus & Ziemann, 2015). For instance, application of the same priming and test theta-burst protocols may reverse the effect of the test protocol, whereas priming with the opposite protocol was shown to increase the effect of a test protocol (Mastroeni et al., 2013; Murakami, Müller-Dahlhaus, Lu, & Ziemann, 2012). Notably, effects of metaplasticity induced by two consecutive NIBS protocols strongly depend (among other factors like stimulation intensity) on the timing between priming and test protocol (Müller-Dahlhaus & Ziemann, 2015), as the subsequent combination of two similar theta-burst protocols may also result in a nonhomeostatic additive effect on corticospinal excitability (e.g., Goldsworthy, Müller-Dahlhaus, Ridding, & Ziemann, 2015). At the behavioral level, it was shown that the capacity of the motor cortex to undergo LTP-like plasticity in response to paired associative stimulation was abolished immediately after motor training (Stefan et al., 2006). These findings are particularly relevant for cognitive neuroscience studies because many training interventions are combined with NIBS protocols to augment the effect of the behavioral intervention, and the principle of homeostatic metaplasticity may
generalize across cortical areas (e.g., Bocci et al., 2014; Gatica Tossi, Stude, Schwenkreis, Tegenthoff, & Dinse, 2013). Yet, the optimal timing between task and stimulation remains unknown for most interventions, and potential effects of metaplasticity are usually ignored. The above-discussed results suggest that metaplasticity may augment, diminish, or even reverse the expected direction of a given NIBS protocol. Although it remains largely unclear how this affects cognitive processes, such effects may contribute to unexpected results or null findings of NIBS studies in the field of cognitive neuroscience.

**CONTROL CONDITIONS—HOW TO DESIGN A VALID NIBS STUDY**

Having outlined the numerous factors potentially interrupting or confounding the chain of causation for a typical cognitive neuroscience NIBS study, we now discuss the most important experimental control conditions. Control conditions are used to control influences that cannot be completely eliminated in NIBS experiments, such as the unintentional costimulation of nontarget brain regions or networks and peripheral sensory structures, the influence of the participants’ expectations and beliefs regarding the NIBS procedure, and the inevitable dependence of task performance on additional nontarget cognitive functions. Importantly, the control conditions included in an experiment determine the conclusions that can be drawn from the results. The typical rationale of NIBS experiments is that, if a given NIBS protocol, applied at a certain time point relative to a task, affects behavioral performance, this is proof of the causal relevance of the targeted area and/or neuronal activity pattern for the cognitive function operationalized by the task. Depending on the hypothesis, several explicit (or implicit) assumptions are made regarding the specificity of the investigated brain–behavior relationship with respect to the targeted anatomical location, temporal window, and/or oscillatory frequency, as well as the affected aspects of the task. For feasibility reasons, it is rarely possible to realize all control conditions in a single study, particularly not in a fully crossed factorial design. However, any specificity claim needs to be explicitly tested with an appropriate control condition.

**Stimulation-free Condition**

Earlier studies often included trials without stimulation (i.e., “no TMS”) as a control condition, sometimes randomly interleaved with effective stimulation to control for carry-over and practice effects (Sandrini et al., 2011). However, this does not control for any unspecified side effects, and the difference between TMS and no TMS will be obvious for the participant. To avoid carry-over effects between trials or conditions, the duration of intertrial intervals and/or TMS trains may be adopted or trials of similar conditions may be grouped in short blocks.

**Sham TMS**

Sham TMS is meant to mimic sensory costimulation and serve as a placebo condition to control for participants’ beliefs and expectations. Different sham TMS approaches have been adopted, including physical separation of the coil from the scalp with a spacer, placing an additional coil with 90° tilt on top of the active coil and selectively discharging the former as a sham condition, slightly tilting a coil on the scalp to avoid stimulation of the underlying brain region, employing a commercially available sham coil, or developing other “realistic” sham conditions. Moving the coil away from the scalp preserves the airborne sound induced by coil charging but induces little or no bone conduction and completely lacks somatosensory costimulation (ter Braack, de Vos, & van Putten, 2015). Likewise, an additional coil on top or tilting the coil such that only the edge touches the head produces roughly comparable auditory inputs compared to active TMS, yet almost no somatosensory inputs because of a lack of peripheral nerve stimulation. The same holds true for conventional sham coils that reduce the effective stimulation intensity by shielding or opposed current flow, which may prevent effective stimulation of the target area but also of peripheral structures. Consequently, participant blinding is hard to achieve with these approaches. This is particularly problematic for within-participant designs often used in cognitive neuroscience studies. To overcome these limitations, “realistic” sham conditions aim to also mimic the somatosensory side effects. State-of-the-art realistic sham conditions thus combine individually adjusted auditory noise masking to attenuate the click sound for both active and sham conditions (Duecker & Sack, 2015), foam padding beneath the coil to attenuate vibration, a TMS coil discharging at an ineffective distance to control for residual auditory input, and individually adjusted electric stimulation of the skin via surface electrodes beneath the TMS coil to mimic peripheral nerve stimulation (Conde et al., 2019). Yet, even if electric stimulation intensity is individually adjusted, experienced participants likely notice the difference between realistic sham and effective stimulation, as the skin sensation of the electric stimulation will be different (e.g., Mennemeier et al., 2009; Rossi et al., 2007). This is particularly problematic when online TMS is applied at higher intensities to areas where the stimulation may yield unpleasant side effects (such as cranial/facial muscle twitches). These side effects can substantially confound the obtained results, as shown by a negative association between working memory performance and individual ratings of the unpleasantness of TMS (Abler et al., 2005). An active control site is thus always preferable. However, the choice of an active control site can be
tricky, especially if complex cognitive functions are studied that are widely distributed across the brain, precluding the use of most well-matched regions as control sites.

**Sham TCS**

In contrast to most TMS studies, TCS studies usually include sham conditions but no active control sites (i.e., control montages). Sham TCS is commonly realized by ramping up the current to target intensity for, for example, 10–30 sec and immediately ramping it down again without any further active stimulation, thus producing some cutaneous (tingling, itching, burning) sensations in the beginning, when they are also strongest for real TDCS before they habituate, supposedly making real and sham TDCS indistinguishable (Gandiga, Hummel, & Cohen, 2006). However, recent evidence shows that, even if low stimulation intensities around 1 mA are used, participant blinding is compromised (Greinacher, Buhöt, Möller, & Learmonth, 2019; Turi et al., 2019). A recently introduced sham TCS protocol combines a multielectrode montage with controlled shunting of currents via a model-based quantification of transcutaneous and transcranial effects, ensuring constant scalp sensations across the whole stimulation procedure and similar sensations relative to effective TCS (Neri et al., 2020). This protocol was suggested to be superior in participant blinding relative to conventional bifocal ramp-up, ramp-down sham protocols and may provide a realistic sham condition for TCS. As for TMS, carefully matched active control montages are the gold standard also for TCS studies, and an additional sham session with the same montage should serve only as low-level baseline. In TDCS studies, polarity specificity can be tested additionally by applying both anodal and cathodal TDCS with respect to the target area of interest.

**Control via Ineffective Stimulation Protocols**

For some NIBS protocols, it is possible to create an ineffective stimulation protocol, which can be applied to the same region without causing the crucial neuronal effects. For example, a change in intertrain duration made the intermediate theta burst stimulation protocol ineffective (Huang et al., 2005), and replacing theta bursts (triplets) with single pulses also produced no aftereffects (Volman, Roelofs, Koch, Verhagen, & Toni, 2011), since 200 pulses at 5 Hz have likely no lasting effect (Peinemann et al., 2004). The paired associative stimulation protocol (Stefan, Kunesch, Cohen, Benecke, & Classen, 2000) becomes ineffective when using imperceptibly for the participant, a random selection of per se ineffective ISIs (Bergmann et al., 2008). In general, a dose reduction can make an ineffective protocol, for example, by reducing stimulation intensity, number of pulses, or total duration of the application. In addition, a change in TMS coil orientation (usually from orthogonal to parallel with respect to the target gyrus) can be sufficient to abolish a specific effect, not only for MEPs evoked in M1 (Mills et al., 1992) but possibly also for other brain regions (Thielscher et al., 2011; Thut, Veniero, et al., 2011). Except for a simple dose reduction, inevitably changing the amount of sensory costimulation and, possibly, the participants’ beliefs regarding the effectiveness of the stimulation, these controls can be considered high-level control conditions matching the effective experimental protocol relatively well, although they come with the risk of residual transcranial effects on the brain.

**Control Tasks (Task Specificity)**

Task specificity implies that, if a cortical area X is relevant for a cognitive process A but not B, then a given NIBS protocol over area X should selectively modulate task A but not B (Miniussi et al., 2013). This requires a control task that differs from the task of interest selectively in the cognitive function of interest but is matched with respect to task difficulty, low-level sensory input, and supporting cognitive functions (e.g., perceptual, attentional, working memory, executive, or motor demands). Without evidence for task specificity, a NIBS protocol may simply interfere with any of those cognitive functions instead, potentially resulting in the same effect on behavioral performance. Importantly, some tasks (e.g., visual attention or orienting tasks) may also be more strongly affected by sensory costimulation.

**Control Regions (Anatomical Specificity)**

Anatomical specificity means that a NIBS protocol only affects task A when applied to area X but not area Y, because only the former causally contributes to the cognitive function of interest. An active TMS control site or TCS electrode montage is thus needed to demonstrate that the observed effects actually depend on the stimulation of a specific brain region and not only of the brain per se or its sensory input structures. Numerous studies have used the vertex as an active control site, with the rationale that auditory and somatosensory inputs should be roughly similar to that of other target sites, but the brain tissue is located deeper under the scalp and thought to mainly contain sensorimotor representations of the lower body, thus not influencing cognitive task performance (Jung, Bungert, Bowtell, & Jackson, 2016; Duecker, de Graaf, Jacobs, & Sack, 2013). However, depending on the target area of interest, differences in lateralization and unpleasantness are a potential issue. For instance, lateralized sham TMS was shown to pull covert spatial attention toward the corresponding side of space, thereby facilitating target detection in this hemisphere (Duecker & Sack, 2013), which can hardly be achieved with vertex stimulation. TMS of pFC and (anterior) temporal cortex can be particularly unpleasant because of costimulation of the facial nerves and muscles, and vertex
stereotaxic neuronavigation system and individual T1-weighted images is highly recommended. In general, TCS studies benefit from optimized montages or computationally optimized multichannel arrangements that may help to focalize the stimulated area and minimize control montages need to be are not sufficient and control montages need to be symmetrical control frequencies below and above the target frequency (Herring et al., 2019; Romei et al., 2016) to define the number of control frequencies or the distance from the frequency of interest (Herrmann et al., 2013). Obviously, the control frequencies should not be engaged in the task of interest. Individual adjustment of the stimulation frequency may be favorable, at least in the alpha band, although it is still unclear whether stimulation is more effective when it matches the eigenfrequency of the brain (Reato, Rahman, Bikson, & Parra, 2013) or is slightly different (Vossen et al., 2015; Helfrich, Knepper, et al., 2014). Notably, TACS can induce rhythmic retinal phosphenes and cutaneous sensations at most frequencies, although with different thresholds (Kanai, Chaieb, Antal, Walsh, & Paulus, 2008), highlighting again that frequency controls alone are not sufficient and control montages need to be employed.

**Further Considerations**

To guarantee accurate coil or electrode placement and maintenance across the experiment and avoid confounds induced by movement of the stimulation device, the use of a stereotaxic neuronavigation system and individual T1-weighted images is highly recommended. In general, TCS studies benefit from optimized montages or computationally optimized multichannel arrangements that may help to focalize the stimulated area and minimize
unwanted peripheral costimulation (e.g., Khatoun et al., 2018). E-field modeling should be employed to estimate the focality of electrode montages or coil placement, minimize the impact of noncortical stimulation, and optimize stimulation efficiency. As many other studies in the field of cognitive neuroscience, NIBS studies often suffer from relatively small and homogeneous samples (i.e., mainly healthy young student volunteers, with sample sizes < 30 participants), which are not representative of the general population. A priori calculations of statistical power may guide sample size selection, and we principally advocate larger sample sizes, but for most NIBS studies, the effect size is unknown beforehand. Whenever possible, NIBS experiments should be conducted as a within-participant design to reduce interindividual variability and confounds based on imperfect randomization of group membership. If control conditions are realized in a between-participant design (e.g., for a post hoc control experiment) the control condition needs to have the same sample size as the experimental condition to rule out statistical power as a confound.

Twelve General Recommendations for Designing Valid NIBS Studies

As a summary, we make the following 12 general recommendations to be considered when designing a NIBS study in cognitive neuroscience (the applicability of these recommendations may vary depending on the specifics of your research question as well as the technical/logistic feasibility in your laboratory):

1. Know your target! Ensure you have identified the stimulation target with a spatial precision appropriate for your NIBS technique of choice. Consider an fMRI-based localizer for target identification if possible.
2. Simulate the E-field! Use individual E-field modeling based on realistic MR-based head models to corroborate that the E-field is maximal at the target site (sensitivity) and as limited to it as possible (specificity).
3. Adjust stimulation intensity! Stimulation intensity should be adjusted individually, taking coil–cortex distance into account, even if no ideal reference value exists yet (% RMT is established). The estimation of induced E-field strength is not (yet) state-of-the-art but recommended.
4. Use neuronavigation! Individual MR-based neuronavigation ensures precise TMS coil/TCS electrode placement and maintenance within and across sessions and is a basic prerequisite for high-quality TMS and high-definition TCS studies.
5. Combine NIBS with neuroimaging! Use neuroimaging (e.g., fMRI, EEG/MEG) to establish proof or target engagement, that is, demonstrate that the desired neuronal effects have been induced in the target region/network and that no unintended activations of other regions/networks can confound the results. This also allows to relate NIBS-induced neuronal and behavioral effects.
6. Include a control site! An active control site (TMS) or montage (TCS), well matched for sensations and annoyance, is a strong control for sensory costimulation confounds and serves to establish anatomical specificity. An additional (realistic) sham condition as baseline is optimal.
7. Include a control task! An appropriate control task, which does not involve the cognitive function of interest but is matched for task difficulty, is a strong control for possible confounding via coaffecting supporting cognitive functions (e.g., attention) and serves to establish task specificity.
8. Include control time windows (if applicable)! If the timing of the target neuronal process matters (e.g., when doing mental chronometry), stimulation at different time points/windows is necessary to establish temporal specificity.
9. Include control frequencies (if applicable)! If you want to demonstrate entrainment or similar frequency-specific effects, neighboring control frequencies (ideally both lower and higher) are mandatory. An arrhythmic control stimulation can control for the number of stimulation pulses/cycles and the sheer presence of rhythmicity but cannot establish frequency specificity.
10. Reduce variability wherever possible! Within-participant designs reduce interindividual variability, but between-participant designs are necessary when repeated stimulation or task performance is problematic for reasons of blinding or learning effects. Internal and external contextual factors (e.g., time of day, arousal) should be kept as comparable as possible between conditions.
11. Prevent carry-over and order effects! For offline NIBS protocols, control conditions need to be conducted in separate sessions several days apart and counterbalanced to ensure that previously induced synaptic plasticity cannot systematically interact with the current protocol.
12. Choose your DV wisely! Thoroughly consider the outcome measure best reflecting the expected change in cognitive function. Titrate task difficulty individually to a level where performance becomes sensitive to even small disturbances of underlying computation.

Conclusion

We have outlined major challenges and potential pitfalls for experimentally testing and interpreting the chain of causation for NIBS studies in cognitive neuroscience. We hope to raise awareness for the potential confounds
and provide a guide for designing valid NIBS experiments. On the basis of the above-discussed studies, a promising avenue for the future will be the multimethod combination of NIBS with computational modeling and neuroimaging to map stimulation-induced changes at the neuronal and network level and link these changes with cognitive and behavioral effects. With respect to computational modeling approaches, particular challenges include the modulation of the dynamics of long-term plastic aftereffects of different NIBS protocols and the transfer from relatively easy decision-making processes to more complex cognitive functions (e.g., language, social cognition, or problem solving). Recent advances in behavioral modeling may further help to bridge the gap between cognitive theories and behavioral outcome measures of NIBS experiments, and the inclusion of more natural tasks or stimuli will increase the ecological validity. To deepen our understanding of the modulatory effects of NIBS protocols, future studies should relate the simulated strength of the induced E-field in the target area to neuroimaging-based assessments of target engagement and behavioral outcome measures. As NIBS studies are complex and time consuming and the modulatory effects are often small and variable across participants, a particular challenge for the future is the inclusion of larger sample sizes to guarantee sufficient experimental power. Here, multicenter approaches may identify valid paradigms for NIBS studies that may provide further insight into causal structure–function relationships.

Acknowledgments
T.O.B. was supported by the Boehringer Ingelheim Foundation. G.H. was supported by the Max Planck Society.
Reprint requests should be sent to Til Ole Bergmann, Leibniz Institute for Resilience Research, Wallstraße 7-9, 55122 Mainz, Germany, or via e-mail: til-ole.bergmann@lir-mainz.de.

REFERENCES
Asamoah, B., Khatoun, A., & McLaughlin, M. (2019). tACS motor system effects can be caused by transcutaneous stimulation of peripheral nerves. Nature Communications, 10, 266. DOI: https://doi.org/10.1038/s41467-018-08183-w. PMID: 30655523. PMID: PMC636776
Beaulieu, L.-D., Flamand, V. H., Massé-Alarie, H., & Schneider, C. (2017). Reliability and minimal detectable change of