Paradoxical Inhibition of Autonomic Output Following Cortical Suppression: When Inhibiting an Inhibitor Results in More (Not Less) Inhibition.

Jason J Braithwaite[1], Giulio Ruffini[2], Haydn Farrelly[1], Ricardo Salvador[2] Hayley Dewe[3], Shalmali D Joshi[1], Derrick G Watson[4], Jean-Philippe Thivierge[5, 6]

[1]Department of Psychology, Lancaster University
[2]Neuroelectrics Barcelona, Spain
[3]Department of Psychology, University of Durham
[4]Department of Psychology, Warwick University
[5]School of Psychology, University of Ottawa, 156 Jean-Jacques Lussier, Ottawa, ON K1N 6N5, Canada,
[6]Brain and Mind Research Institute, University of Ottawa, 451 Smyth Rd., Ottawa, ON K1H 8M5, Canada

Corresponding author: j.j.braithwaite@lancaster.ac.uk

Abstract

Paradoxical effects in brain stimulation, where inhibition of a putative inhibitory region leads to stronger (not weaker) downstream suppression, challenge overly simplistic classical assumptions about excitation-inhibition interactions. Joshi et al. (2024) demonstrated that optimised inhibitory multi-channel transcranial direct current stimulation (MtDCS) over the right ventrolateral prefrontal cortex (rVLPFC) enhanced suppression of autonomic skin conductance responses (SCRs). However, this only occurred in individuals with signs of elevated cortical hyperexcitability and only in the cathodal condition. This inhibitory effect challenges simplistic polarity-based single-node functional models that assume fixed excitatory or inhibitory outcomes. We propose that such *paradoxical inhibition* may be understood through systems-level (network) accounts such as precision predictive coding and cortical gain control. We further examine how microcircuit models such as Inhibition-Stabilised Networks (ISNs) might be scaled up to explain these paradoxical outcomes. Together, these frameworks suggest that suppression of excitatory drive, under conditions of heightened salience and trait vulnerability, may paradoxically trigger enhanced inhibitory output as a stabilizing response. This challenges simple cause-effect models of neuromodulation and highlights the need for theories to incorporate emergent network dynamics. As a novel entry point into paradoxical dynamics, these findings require careful replication and extensions to probe how suppressed excitation may trigger stabilising inhibition across distributed neural systems.

Background

The human brain is composed of vast, recurrently connected networks of excitatory (E) and inhibitory (I) neurons that work together to support stable processing (Haider et al., 2006; Litwin-Kumar & Doiron, 2012). Excitatory pyramidal neurons project across cortical layers and between brain areas, forming the primary substrate for feed-forward and feedback signalling. Inhibitory interneurons play a critical role in shaping timing, cortical gain, and preventing excessive excitation within neural systems (Denève & Machens, 2016; Douglas & Martin, 2004; Shadlen & Newsome, 1994). This dynamic interplay allows the cortex to balance responsiveness and stability across a wide range of sensory, cognitive, and emotional tasks.

A fundamental outcome of these dynamic interactions is the balance of excitation and inhibition within a given system (E/I balance). This balance is not static; it is in a constant state of flux, where opposing E and I processes collide. Intuitively, an excess of excitation can trigger runaway neuronal activity, whereas excessive inhibition can suppress responses to sensory input. Even subtle disruptions in this equilibrium can lead to significant changes resulting in cortical dysfunction and aberrant experience (Anderson et al., 2000; ffytche, 2008; ffytche et al., 1998; Froemke, 2015; Rubenstein & Merzenich, 2003; Yizhar et al., 2011). Maintaining the appropriate ratio of excitation to inhibition is therefore essential for cortical circuits to function within a dynamically stable yet responsive regime.

Previous work has demonstrated that optimised multi-channel transcranial direct current stimulation (MtDCS), configured to deliver inhibitory current to the right ventrolateral prefrontal cortex (rVLPFC), produced a paradoxical inhibitory effect on downstream autonomic responses (Joshi et al., 2024). In that study, Joshi et al. examined whether targeted cathodal / inhibitory MtDCS over the rVLPFC would modulate skin

conductance responses (SCRs) to aversive body-threat stimuli. The rVLPFC is theoretically relevant because it forms part of a broader cortico-limbic-autonomic regulatory network with functional relationships to the amygdala, insular cortex, anterior cingulate cortex, and downstream sympathetic pathways (Beissner et al., 2013; Cheng et al., 2022; Craig, 2009; Critchley, 2002; Medford, 2012; Sierra & David, 2011). Within this network, the amygdala and insula are particularly important for threat salience, interoceptive awareness, autonomic arousal, and the mediation of cognitive-affective states, while SCRs provide an index of sympathetic output (Boucsein, 2012; Craig, 2002; 2003; 2009; Dawson et al., 2007).

Based on previous work in dissociative depersonalisation (Jay et al., 2014; 2016; Sierra & Berrios, 1998; Sierra & David, 2011; Sierra et al., 2002; Phillips et al., 2003) and broader evidence implicating the rVLPFC in inhibitory / modulatory control over affective-autonomic processing (Cheng et al., 2022; Critchley et al., 2000; Critchley, 2002; Jay et al., 2014; Nagai et al., 2004; Phillips et al., 2001; Sierra & David, 2011; see Millman et al., 2024 for a recent review), the classical prediction was that inhibiting the rVLPFC would reduce its regulatory influence on downstream affective-autonomic targets thus releasing them from aberrant inhibition.

Contrary to this prediction, it was found that greater suppression of SCRs occurred. Critically, this effect was not generic: it was observed selectively in individuals with elevated trait-based signs of cortical hyperexcitability. The effect was not evident in the same way for individuals with lower trait-based cortical hyperexcitability, and there were no effects following anodal stimulation. This pattern suggests that the observed suppression was not simply a consequence of stimulation polarity or reversal, but reflected an interaction between prefrontal perturbation, trait-level cortical gain, and stimulus-driven salience.

This apparent *'paradoxical inhibition'* (where inhibiting an inhibitory brain system led to more, not less, downstream inhibition) highlights a fascinating interplay between baseline brain states, brain stimulation, and network-level inhibitory control. It also challenges conventional assumptions about how brain stimulation can influence neural systems, particularly the idea that cathodal stimulation simply down-regulates activity in a target area. Instead, it points to a more complex, non-linear interaction between stimulation parameters, underlying network excitability, trait vulnerability, and stimulus context.

One plausible explanation for this seemingly paradoxical effect is that when excitable (E) neural systems are externally suppressed, downstream systems may interpret this new weakened or ambiguous input as destabilising, and compensate by increasing inhibitory output to maintain stability. Homeostatic plasticity mechanisms contribute to this process by acting to restore firing rates toward target levels, thereby re-establishing balance across circuits and stabilizing their patterns of activity (see Godin et al., 2025; for recent evidence from computational modelling). This is akin to general adaptation effects in which an external input may become reduced by systems designed to reach a standard equilibrium point (Hengen et al., 2013).

Although this discussion uses Joshi et al. (2024) as the primary empirical anchor, polarity-atypical and counterintuitive outcomes have been reported more broadly across non-invasive brain stimulation research. These collective findings establish an important precedent: stimulation polarity does not map reliably onto simple functional outcomes in a fixed anodal-excitatory / cathodal-inhibitory manner (i.e., atypical polarity effects can and do occur). Particularly relevant examples include work showing polarity and task-context dependent modulation of right inferior frontal networks (Li et al., 2019), baseline-dependent effects of dorso-lateral pre-frontal cortex (DLPFC), stimulation on visual working memory

(Hsu et al., 2016), and individual-difference dependent effects of rVLPFC stimulation on inhibitory control (Bell et al., 2022). These studies are not direct examples of paradoxical inhibition, but they are important because they show that polarity-atypical outcomes can emerge in frontal control systems, can depend on baseline performance or participant characteristics, and can involve regions close to the rVLPFC circuitry considered here.

Furthermore, cathodal stimulation, typically assumed to be inhibitory, has been shown to facilitate performance in both motor and prefrontal regions (Antal et al., 2004; Jacobson et al., 2012; Garnett et al., 2015; Wyss et al., 2024). Other studies have reported polarity-independent improvements, where both anodal and cathodal stimulation produced comparable behavioural enhancements, consistent with nonlinear effects that cannot be explained by polarity alone (e.g., DeLaRosa et al., 2020). Finally, several studies have produced null or non-differential effects even in regions where polarity-driven responses are often anticipated, including the dorsal anterior cingulate cortex, cerebellum, and primary motor cortex (Gorrino et al., 2023; Maldonado et al., 2019; Cabibel et al., 2018). Collectively, these findings challenge simple polarity-based accounts of neuromodulation and point to a more complex landscape in which neurostimulation effects are shaped by target anatomy, montage configuration, baseline excitability, task context, and network state.

It is important, however, not to conflate polarity-atypical effects with the narrower phenomenon described here as Paradoxical Inhibition. Polarity reversals typically refer to outcomes that run counter to simple anodal-excitatory / cathodal-inhibitory expectations, such as cathodal facilitation or anodal impairment. Paradoxical inhibition, as used here, refers to a more specific pattern: suppression of excitatory drive within a regulatory cortical region being associated with greater downstream inhibition of autonomic output under defined trait- and state-dependent conditions. Thus, polarity-atypical findings provide useful evidence

against simple linear polarity models, but they do not by themselves explain the particular pattern observed following cathodal rVLPFC stimulation in high-gain individuals viewing aversive stimuli. Importantly, many of these studies used conventional (bipolar) or fixed montage approaches rather than optimised, computer-modelled multi-channel approaches based on realistic head models and therefore provide an important precedent for polarity-atypical effects without directly addressing the targeted network-level mechanism considered here.

In the remainder of this article we examine paradoxical findings (with a focus on our previous work, though with relevance to broader findings) via three complementary theoretical models; (i) precision predictive coding, (ii) cortical gain control, and (iii) the recently emerging field of inhibition-stabilized networks (ISNs), each offer distinct but overlapping insights into how the suppression of excitation can lead, paradoxically, to enhanced inhibition.

Advances in Brain Stimulation: Optimised Multi-Channel Direct-Current Stimulation (MtDCS)

Recent advances in transcranial electrical stimulation (tES) have dramatically improved our ability to target brain regions with greater anatomical and functional precision. Although widely promoted as an upgrade over traditional bipolar montages, 'high-definition' (HD) 4×1 configurations (with one active and four return / reference electrodes arranged in a ring: Borckardt et al., 2012; Datta et al., 2009; Edwards et al., 2013; Kuo et al., 2013; Villamar et al., 2013) remain fundamentally limited in their capacity to accurately target anatomically complex regions such as the right ventrolateral prefrontal cortex (rVLPFC).

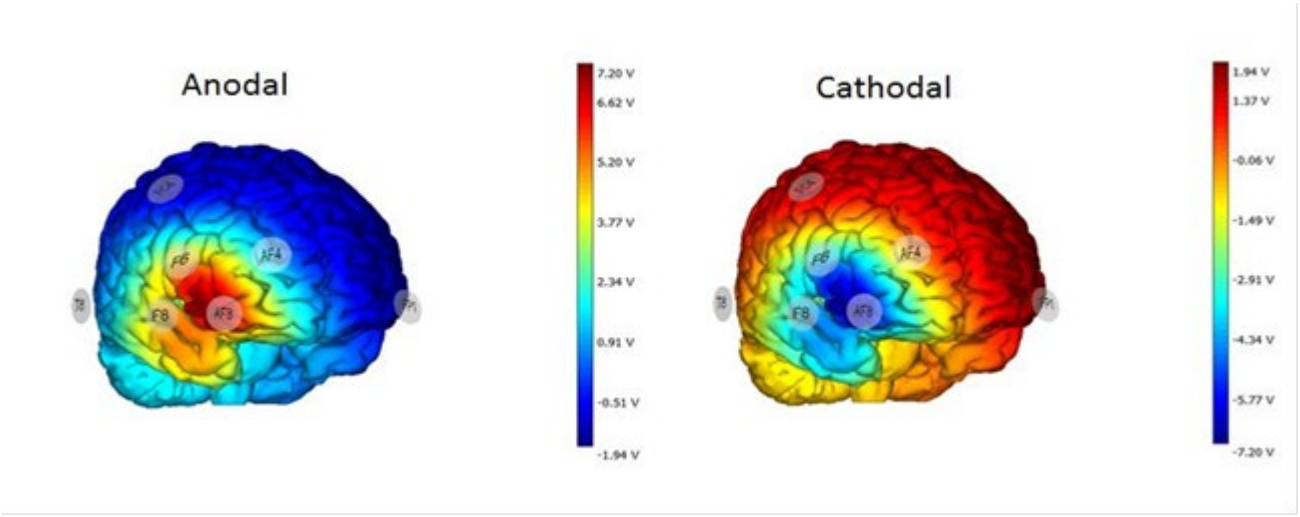
The rVLPFC is a spatially extended, L-shaped structure that spans both lateral and ventral surfaces, incorporating parts of the inferior frontal gyrus and anterior insula. This geometry renders it poorly suited to focal 4×1 or standard bipolar montages, which lack the spatial coverage needed to engage the full extent of this region. Even studies that use basic modelling often rely on idealized head models and oversimplified assumptions about current flow (Datta et al., 2011; Edwards et al., 2013; Kuo et al., 2013). As well as being constrained into using only 5 electrodes (which may not be the optimal solution), such approaches may fail to meaningfully modulate the functional area of interest.

By contrast, optimised multi-channel tDCS (MtDCS), utilising finite element models (FEM) based on realistic MRI-derived head models, allows for more precise current steering and targeting across anatomical locations. This technique uses multiple small electrodes, with individualised current weights and polarities (informed by tissue-specific conductivity) calculated via FEM-based modelling of realistic head anatomies. It enables the electric current to be steered and shaped across cortical surfaces based on real-world conductivity values, anatomical curvature, and intended functional engagement (Ruffini et al., 2014; 2018; 2024; Miranda et al., 2013; 2018; Salehinejad et al., 2026; Salvador et al., 2010, Cheng et al., 2026). These montages can be configured to deliver either excitatory (anodal) or inhibitory (cathodal) stimulation across multiple regions more precisely localised between electrodes. This is achieved through the control of the normal component of the electric field to the cortical surface, based on the "lambda-E" model of polarization of pyramidal cells (Ruffini et al., 2014, Galan-Gadea et al., 2023)[1].

Figure 1

[1] Note, here, 'inhibitory' refers to the intended and modelled suppressive effect of the cathodal MtDCS montage on excitatory drive within the targeted rVLPFC region. It does not imply a simple one-to-one mapping between stimulation polarity and downstream autonomic outcome; indeed, the central claim of the present paper is that such network effects may be nonlinear and state-dependent.

Optimised 7-channel MtDCS montages targeting the rVLPFC, adapted from Joshi et al. (2024). Left: Anodal configuration with inward current (red); Right: Cathodal configuration with outward flow. Montages were modelled using 3D finite element methods to optimize electrode placement, polarity, and weighting (Neuroelectrics, Barcelona, Spain).

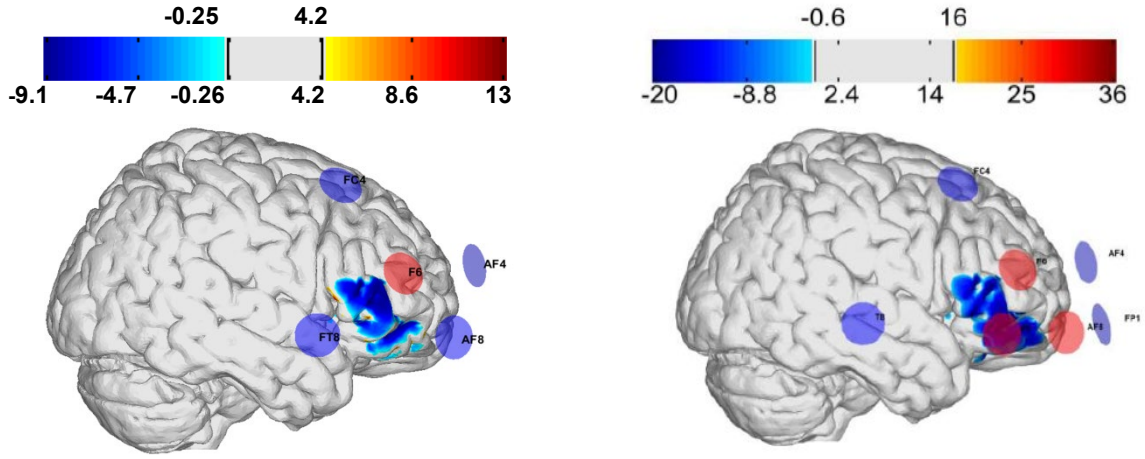


Such FEM-based optimisation has confirmed that standard 4×1 configurations yielded sub-optimal engagement with the rVLPFC, reflected in low ERNI values (Error with Respect to No Intervention = -1841mV^2/m^2). In contrast, the 7-channel optimised montage achieved a far superior ERNI scores (ERNI = -4245mV^2/m^2), with normal electric fields closely aligned to the shape and extent of the rVLPFC (see Figures 1 and 2)[2]. These optimised montages ensured meaningful engagement with both lateral and ventral components of the region, an outcome that standardised (bipolar and 4x1) configurations cannot achieve.

Figure 2

[2] Note, the larger the negative ERNI value, the better the 'fit' between the intended stimulation region and what can be achieved by the montage. ERNI is ostensible a measure of 'fit' or 'coverage'.

Electric field modelling for a standard 4×1 HD-tDCS montage (left) versus a model-optimised 7-channel configuration targeting the rVLPFC (right). The 4×1 montage shows limited coverage, while the 7-channel montage delivers more spatially precise fields aligned with the region's morphology. Simulations used realistic head models via the Neuroelectrics platform.



In sum, the use of FEM-optimised MtDCS represents a significant methodological advance, allowing researchers to excite or suppress anatomically complex, functionally specific brain regions with a level of precision that is not attainable with standard bipolar or 4×1 HD-tDCS montages. Given the central role of rVLPFC in inhibitory control, such precision is essential when interpreting downstream cognitive-behavioural or autonomic effects.

Frameworks (I and II): Precision-Weighted Predictive Coding, Gain Control & Paradoxical Effects

Two influential and related systems-level frameworks are: (i) predictive coding and (ii) cortical gain theories, with the former focusing on hierarchical inference and the latter on signal regulation in the brain (Ferguson & Cardin, 2020; Friston, 2005; 2008; Rao & Ballard, 1999). Indeed, as the theory of predictive coding has evolved it has assimilated the notion of cortical gain control into its predictive process and consequently is now often referred to as precision-weighted predictive coding (Aukstulewicz & Friston, 2016; Feldman & Friston, 2010; Kanai et al., 2015; Spratling, 2008; see also Millidge et al., 2022). Both frameworks converge on a key idea: that the brain dynamically adjusts the weight and amplitude of signals based on their inferred reliability (precision) or salience.

In predictive coding, neural hierarchies minimise prediction error by comparing incoming sensory data against pre-existing top-down predictions or 'priors' (Friston, 2005; Clark, 2013; Seth, 2013). Crucially, these error signals are not all treated equally: their influence is scaled by an estimate of precision, a computational proxy for reliability (Kanai et al., 2015). This weighting is thought to be implemented biologically, via gain control of prediction error and through inhibitory interneurons that modulate their output (Barrett & Simmons, 2015; Bastos et al., 2012; Etkin et al., 2011; Feldman & Friston, 2010; Haider, & McCormick, 2009; Kanai et al., 2015).

In this framework, an MtDCS driven reduction in excitatory drive (E↓) within the rVLPFC, might diminish the regions capacity to amplify salient input, effectively flattening localised gain control and impacting precision. As a result, downstream systems that depend on rVLPFC modulation may interpret this weakened input as unstable triggering broader network-level inhibitory processes to intervene and restore the (E/I) balance. According to

this account, paradoxical inhibition of biophysical markers like skin conductance responses may emerge not from a direct increase in inhibition, but from degraded excitatory signalling (due to suppression of (E↓), that is subsequently over-suppressed by downstream stabilising mechanisms (Friston & Kiebel, 2009; Spratling, 2010).

This dynamic fits closely with trait and state-dependent findings because high cortical hyperexcitability (high tonic gain) may make the system more sensitive to small perturbations, resulting in an overly-strong inhibitory response when excitatory drive is diminished (Filmer et al., 2019; Godin et al., 2025; Joshi et al., 2024; Vergallito et al., 2023; Vogels & Abbott, 2009). In terms of neuronal transfer functions, this corresponds to being positioned on the steep part of a nonlinear (sigmoid-like) input–output curve, where small changes in input can produce disproportionately large changes in output, reflecting intrinsic differences in current-to-firing rate (I-F) sensitivity across neurons, that is, how strongly a neuron's firing rate increases for each increment of input current.

While predictive coding emphasises hierarchical inference and precision-weighted signalling, gain control models address the biophysical regulation and tuning of cortical responsiveness (Abbott et al., 1997; Aston-Jones & Cohen, 2005; Haider & McCormick, 2009; Servan-Schreiber et al., 1990). Gain control describes how neural systems adjust response amplitude in a nonlinear, stabilising manner (Carandini & Heeger, 2012; Chance et al., 2002; Ferguson & Cardin, 2020; Rabinowitz et al., 2015). Gain control allows the cortex to adjust signal amplification according to context, selectively enhancing behaviourally relevant inputs while dampening noisy or destabilising signals, a process that arises from finely tuned interactions between dynamic excitatory and inhibitory mechanisms (Ferguson & Cardin, 2020; Harris & Thiele, 2011; Li et al., 2020; Vogels & Abbott 2009).

When the input strength deviates from the system's preferred operating range, either too strong (as in hyperexcitable states) or too weak (as with degraded cortical drive), inhibitory engagement scales to rein the system's output back toward an optimal range (Abbott & Chance, 2005; Abbott et al., 1997; Chance et al., 2002; Schwartz & Simoncelli, 2001; Zhang & Abbott, 2000). In the case of findings like Joshi et al., (2024), cathodal MtDCS reduced excitatory drive from rVLPFC pyramidal neurons (E↓). In individuals with high trait hyperexcitability viewing emotionally aversive stimuli, the input feeding into the amygdala-insula network would deviate from the range of expected amplitude.

This deviation from expected gain could trigger a non-linear inhibitory overshoot within the amygdala-insula circuitry, resulting in a paradoxical suppression of affective and autonomic responses (Muir & Kittler, 2018; Schwartz & Simoncelli, 2001). Unlike predictive coding, this framework involves no interpretive inference. Instead, it reflects a biophysical normalisation process likely mediated by complex interneuron cascades operating over different time scales. In this view, overshoot occurs because inhibition ramps up as a distributed compensatory response, but due to its slower time course it continues increasing beyond what is needed, thereby suppressing responses more than intended. The suppression of SCRs may therefore reflect a response to unstable input gain, rather than a direct reaction to the signal's predicted content.

Although framed differently, both predictive coding and gain control provide a coherent systems-level explanation for how suppressed excitatory output from a neural system could, paradoxically, result in enhanced downstream inhibition, particularly in emotionally salient contexts and in individuals with heightened baseline excitability. They suggest that the brain's regulatory systems are not simply driven by the local polarity or amplitude of brain stimulation, but by inferences about endogenous stability, the excitatory-

inhibitory balance, and expected or predicted levels of activation, salience, and uncertainty. This makes them potentially powerful candidate frameworks for understanding paradoxical effects at the systems level.

A drop in cortical excitatory output may not merely reduce signal strength; it may impact the perceived reliability of that signal, prompting stabilising inhibitory systems to increase their participation in the process. This positions gain modulation not just as an amplitude control mechanism, but as a functional bridge between network instability and stabilising control. Broader findings from brain stimulation studies (e.g., Antal et al., 2004; Jacobson et al., 2012; Benwell et al., 2015; Iwabuchi et al., 2017; Joshi et al., 2024) could be taken to implicate gain control as a biologically plausible, systems-level mechanism through which nonlinear unexpected responses (including paradoxical inhibition), may emerge.

Frameworks (III): Paradoxical Inhibition and Inhibition Stabilised Networks (ISNs). A Conceptual Scaffold?

A third framework is that of inhibition-stabilised networks (ISNs), where strong recurrent excitation is dynamically balanced by strong localised inhibitory feedback (Ozeki et al., 2009; Tsodyks et al., 1997; see also, Godin et al., 2025; Isaacson & Scanziani, 2011; Rubin et al., 2015; Sanzeni et al., 2020; Soldado-Magraner et al., 2022; Wu & Gjorgjieva, 2023).

A key signature of Inhibition-Stabilized Networks (ISNs) is the paradoxical inhibitory response, where direct stimulation of inhibitory (I) interneurons initially increases their firing (as expected), which in turn, then suppresses the excitatory (E) population via feedback inhibition. As inhibitory neuron activity in an ISN is largely driven by recurrent excitation from the (E) population, this suppression of (E) reduces the excitatory drive back onto the (I)

neurons. Consequently, inhibitory activity paradoxically falls despite the continued external direct stimulation onto (I) neurons. This phenomenon highlights the non-linear interdependence of (E) and (I) populations under conditions of strong recurrent coupling.

At the microcircuit level, ISNs describe a regime where inhibitory interneurons are primarily driven by recurrent excitation from pyramidal cells (Tsodyks et al., 1997; Sanzeni et al., 2020). Consequently, increasing inhibitory drive initially elevates interneuron firing, but the resulting suppression of excitatory activity reduces the recurrent input onto the inhibitory population itself, leading to a paradoxical decrease in inhibitory firing rates. Importantly, this reduction in inhibitory activity does not imply reduced net inhibition onto pyramidal neurons; rather, the increase in inhibitory drive can produce a net increase in inhibitory influence on the excitatory population. This dissociation between interneuron firing rates and the net inhibitory effect on pyramidal cells highlights the nonlinear behaviour of ISN microcircuit dynamics, where population-level inhibition is governed by recurrent interactions rather than by the activity of inhibitory neurons alone (Godin et al., 2025).

Based on recent computational work, mechanisms associated with ISNs at the microcircuit level can be extended to macroscale dynamics. Neural mass models such as coupled excitatory-inhibitory Wilson-Cowan populations demonstrate that paradoxical ISN-like responses can emerge in larger-scale networks (Godin et al., 2025). In these models, strengthening inhibitory pathways suppresses excitatory activity, which in turn reduces the recurrent drive that sustains inhibition, leading to a paradoxical decrease in inhibitory activity. Animal studies support this framework, showing that optogenetic manipulation of inhibitory circuits can propagate through distributed networks and produce counterintuitive changes in population dynamics (Sanzeni et al., 2020). Importantly, extending ISN principles to the macroscale requires shifting from a view of inhibition based on specific localised

cellular interactions to one where inhibitory effects can emerge from distributed feedback loops spanning interconnected regions. Within this framework, inhibition can be amplified or attenuated as a consequence of the network's tendency to restore dynamical stability.

This perspective is particularly relevant for understanding cortical hyperexcitability in high-gain regimes. In such contexts, network activity operates near an unstable operating point where small perturbations in excitatory drive can elicit disproportionately large compensatory responses. Reducing excitation in this context does not merely lower activity. Rather, it perturbs the balance of recurrent interactions that maintain stability. As excitatory input to inhibitory pathways is diminished, the regulation of inhibition becomes disrupted, and network-level feedback can transiently amplify inhibition to promote stability. Thus, a decrease in excitation can paradoxically strengthen inhibition, reflecting compensatory mechanisms to re-establish stable dynamics.With the viewpoint that ISNs may be scaled from the microscopic-level to macroscopic-level circuits, their effect may conceptually scale-up to broader regional dynamics in systems like the rVLPFC, which exhibits strong feedback inhibition and such influences affective regulation via functional interactions with amygdala-insula circuitry (Tsodyks et al., 1997). This *'ISN-like'* framework could also explain how inhibitory stimulation of the rVLPFC could lead to broader suppression of autonomic affective responses (SCRs). ISNs rely on a strong inhibitory coupling and fast-acting interneurons are seen as critical to the process at the cellular level (Parvalbumin-positive (PV+) interneurons: Hu et al., 2014; Kawaguchi & Kubota, 1997; Tremblay, Lee, & Rudy, 2016). However, the typical view of PV+ interneurons is that they are likely too fast to be involved in the suppression of slower-acting autonomic SCRs.

The rVLPFC and the amygdala-insula network also contain an abundance of slower-acting classes of inhibitory interneurons also capable of more sustained inhibition (including

somatostatin (SOM+) and neurogliaform-like (NGF) interneurons: Capogna & Pearce, 2011; Silberberg & Markram, 2007; Tamas et al., 2003; Tremblay et al., 2016; Urban-Ciecko & Barth, 2016; Wilson et al., 2012). One speculative mechanism is that these later-acting interneurons could sustain inhibition of autonomic responses over a more prolonged time period via their own direct effects onto pyramidal neurons, or by the reduction of their own inhibitory influence on PV+ interneurons. Therefore, the concept of a fast-acting trigger and a slower acting suppression are, at the very least, neurologically plausible at the cellular level.

Furthermore, inhibitory interneurons may not need to sustain inhibitory activity throughout the full duration of an SCR in order to modulate SCR strength. Rather, what might be critical is a sufficiently strong inhibitory influence during the early phase of stimulus appraisal and autonomic gating, perhaps within the first 0-1.5 seconds or so. Suppression during this window may be enough to blunt the peak output of the system and prevent the full expression of autonomic responses. This is consistent with observations where SCRs were not eliminated, but were reliably reduced in strength, suggesting early inhibitory gating rather than full autonomic shutdown (Joshi et al., 2024).

Three further observations are worth highlighting. First, as a nonlinear response, ISN dynamics do not require large perturbations to produce strong effects. Modest changes can trigger disproportionately large inhibitory responses when the network is near a critical and unstable regime (Sanzeni et al., 2020; Tsodyks et al., 1997). Second, ISNs do require that the perturbation engages a sufficiently broad proportion of the inhibitory interneuron population. If only a narrow subset of interneurons is recruited, paradoxical effects may not be elicited and the network would remain in a classical regime (Ozeki et al., 2009; Rubin et al., 2015). Finally, these points suggest that MtDCS, may strike an important balance in terms of network coverage and weak, though sufficient, perturbation and may therefore be better

suited for revealing such nonlinear systems-level effects than more traditional stimulation approaches that are: (i) poorly steered into the critical area and lack coverage or (ii) are too focal or too intense, both of which risk bypassing these network dynamics altogether[3].

The role of cortical hyperexcitability may also dovetail neatly with ISN conceptions, as such an increase in background traits or 'tone' could be interpreted by the system as increased gain or instability, priming the network more towards a nonlinear regime. More formally, recent models suggest that homeostatic plasticity can shift networks between ISN and non-ISN regimes as a function of tonic input, supporting the view that hyperexcitable systems may lie closer to an ISN threshold, where modest perturbations elicit strong nonlinear inhibitory responses (Godin et al., 2025; Soldado-Magraner et al., 2022). Put simply, cortical hyperexcitability may increase the effective gain of the (I) response in some systems not by increasing (I) directly, but by placing the system in a regime where (I) becomes extremely reactive to small changes and perturbations in (E).

If cortical hyperexcitability shifts networks toward an ISN-like regime, then similar non-linear E-I interactions should be observable in well-characterized circuits governing affective and autonomic control. In this context, rodent work on medial prefrontal-amygdala-brainstem pathways provides a useful bridge between ISN-like circuit dynamics and the proposed macroscale account. In fear-conditioning paradigms, prelimbic cortex is typically associated with fear expression via excitatory projections engaging basolateral and central amygdala circuits, whereas infralimbic cortex is more closely linked to fear suppression and extinction through recruitment of inhibitory amygdala microcircuits that dampen central amygdala output to hypothalamic and brainstem targets (Sierra-Mercado et al., 2011).

[3] Though this remains speculative at the current time.

Crucially, these pathways operate in state-dependent regimes governed by balanced excitation and inhibition (Arruda-Carvalho & Clem, 2014). As such, they provide a useful animal analogue for the present proposal, where reducing prefrontal excitatory drive under a high-gain state may perturb the prefrontal-amygdala-brainstem circuit and engage compensatory inhibition within the amygdala / insula complex. Hence, macroscopic ISN mechanisms may characterize the dynamics of broadly distributed circuits involving the interaction of excitatory and inhibitory pathways, including cortex-amygdala-brainstem connectivity where changes in excitatory drive may lead to the suppression of fear-related responses.

While conceptually promising, it is important to acknowledge several key caveats regarding the application of ISN theory. ISNs have primarily been observed at the microcircuit level. However, the idea of extending ISN dynamics to larger systems represents a plausible extrapolation of mechanisms established at the microscale level. The networks necessary to support nonlinear ISN regimes are ubiquitous across cortex and limbic structures, and although no studies have yet demonstrated ISNs at the full macroscopic scale discussed here, several theoretical accounts (Eskikand et al., 2023; Herstel et al., 2021; Soldado-Magraner et al., 2022) and empirical findings (Eskikand et al., 2023) already point toward ISN-like effects at larger scales than originally observed. In this sense, the notion of scale is an open theoretical step forward.

Moreover, it is important to note that the classic ISN effect typically emerges from the direct activation of inhibitory interneurons (I), whereas the effects reported by non-invasive brain stimulation involve reducing activity in excitatory systems (E). This paradoxical reduction in excitatory activity remains to be systematically investigated in theoretical models of ‘ISN-like’ processes.

It is also theoretically possible that a paradoxical increase in inhibitory activity could emerge from interactions between fast and slow inhibitory populations, perhaps without invoking notions of a classical ISN regime. For example, if a reduction in excitatory drive decreases feed-forward activation of a slower inhibitory population (SOM+ / NPY / NGF), and this slower population normally suppresses a faster inhibitory population, then its diminished output could lead to disinhibition of the fast (PV+) pathway. Due to the slower action of this second population, its delayed response to reduced excitation may fail to suppress the fast interneurons effectively, resulting in a transient but paradoxical increase in fast inhibition. Because the slower inhibitory population exhibits a lagged response to the decrease in feed-forward excitation, its ability to inhibit the faster inhibitory population would be diminished, possibly resulting in a significant net increase in fast inhibition. Teasing apart these possibilities from potential ISN-like effects will be an important endeavour for future investigations.

Table 1

Summary of Predictive Coding, Gain Control, and ISN mechanisms.

Framework	Gain modulation	Paradoxical inhibition	State dependence (hyperexcitability)	Microcircuit implementation	Systems-level interpretation	Explanation of Paradoxical Inhibition	Key References
Predictive Coding	✓	✓	✓	X	✓	Degraded excitatory drive interpreted as unreliable.	Auksztulewicz & Friston (2016); Barrett & Simmons (2015); Bastos et al. (2012); Clark (2013); Feldman & Friston (2010); Friston & Kiebel (2009); Rao & Ballard (1999)
Gain Control	✓	✓	✓	✓	X	Non-linear inhibitory overshoot.	Abbott & Chance (2005); Carandini & Heeger (2012); Chance et al. (2002); Ferguson & Cardin (2020)
ISN	✓	✓	✓	✓	X	Suppression of excitatory feedback.	Godin et al. (2025); Sanzeni et al. (2020); Tsodyks et al. (1997); Soldado-Magraner et al. (2022)

Table 1 provides an overview of predictive coding, gain control, and ISN theories and their relevance to systems-level paradoxical effects. Each theory succeeds in explaining key experimental findings on gain modulation, paradoxical inhibition, and state dependence, though through distinct mechanisms. Importantly, these frameworks differ in their level of explanation: predictive coding operates at the systems level, whereas gain control and ISN focus more on microcircuit dynamics. This contrast points to a theoretical gap, namely, the absence of a systems-level framework grounded in gain-mediated ISN principles, which could account for paradoxical effects at a macroscopic scale.

ISN-like logic, and the broader notion of paradoxical effects, can serve as a conceptual bridge between biologically grounded circuit principles and the systems-level operations emphasised by predictive coding and gain-control frameworks. The latter excel at explaining large-scale, context-sensitive behaviour but are under-specified about cellular implementation.

In the present discussion, ISNs act more as a ‘theoretical scaffold’ where we refer to ‘ISN-like’ behaviour at the systems level rather than claiming a canonical microcircuit ISN

within rVLPFC in humans. The aim is to borrow the logic of a paradox, though in terms of brain stimulation findings here, that (E↓) can lead to a net increase in (I↑). By this account, cathodal MtDCS suppresses excitatory output from rVLPFC (E↓). The paradox is expressed downstream: reduced, low-precision prefrontal input propagates to limbic targets (amygdala-insula), where inhibitory gating overshoots, yielding reduced autonomic output (suppressed SCRs). Thus, the measurable inhibitory effect need not be a property of the rVLPFC alone but of the distributed network preceded by prefrontal perturbation, a view fully consistent with precision-weighted predictive coding and gain-control accounts in which poor calibrated precision leads to down-weighting and over-suppression of unreliable signals (see Table 2).

Accordingly, neither our findings nor other paradoxical outcomes in non-invasive brain stimulation confirm a canonical ISN regime. Rather, they reveal paradoxical inhibition following cortical suppression at the intersection of local circuit dynamics and systems-level interpretation. These possibilities warrant careful replication and targeted tests to determine when reduced excitation functions as a catalyst for emergent stabilising inhibition across distributed networks. Explaining such effects will require models that are computationally realistic yet tractable, biologically grounded, and explicitly state-dependent, a synthesis that can sharpen our understanding of neuromodulation and inhibitory control in the human brain.

Importantly, such effects may not have been observed in earlier prefrontal stimulation studies as many were not designed to detect this specific combination of conditions. Most, (i) did not stratify participants according to trait cortical hyperexcitability or related indices of baseline gain, (ii) did not include SCRs as a comparatively selective measure of sympathetic autonomic output, and, (iii) often used conventional bipolar montages in which current flow is relatively diffuse and strongly influenced by the position of the return electrode. Such designs are useful for testing broad polarity-based hypotheses, but they are less well suited to

detecting targeted trait and stimulus-dependent network effects involving the rVLPFC and downstream affective-autonomic circuitry. Thus, the apparent rarity of paradoxical inhibition in the existing prefrontal stimulation literature should not be taken as evidence that the effect is implausible; it may instead reflect the fact that most studies have not combined anatomically realistic and optimised rVLPFC targeting, trait-based stratification, aversive / high-salience stimuli, and sympathetic autonomic measurement within the same design.

Table 2

Mechanistic accounts evaluated for paradoxical SCR suppression under cathodal rVLPFC MtDCS, with predictions and tests.

Level	Mechanism	Paradox account (E ↑ → I ↓)	Signatures / Timescale	Boundary Conditions	Discriminating Tests	Key References
Local	ISN & interneuron hierarchy (PV+, SOM+/NGF)	Suppressed rVLPFC E causes ISN-like or disinhibitory I overshoot downstream, blunting SCR peaks	Non-linear/biphasic E–I; early gating (1–1.5 s); broad I recruitment needed	Near-ISN regime; trait hyperexcitability	TMS–EEG to separate fast/slow inhibition; montage breadth/intensity sweeps; stabilised supralinear fits	Tsodyks et al. (1997); Ozeki et al. (2009); Rubin et al. (2015); Sanzeni et al. (2020)
Network	Gain control / normalization	Input falls outside operating range; normalization increases I to re-center output, over-suppressing autonomic response	Nonlinear IO curves; rapid divisive scaling	High salience; high tonic gain; background synaptic drive	Manipulate amplitude (not uncertainty); relate effect to background drive; pupil-indexed NE	Abbot et al. (1997); Carandini & Heeger (2012); Ferguson & Cardin (2020)
Systems	Precision-weighted predictive coding	rVLPFC E↓ lowers precision of prefrontal signals; downstream circuits downweight/over-suppress unreliable inputs	Effects scale with uncertainty; EC changes; rapid precision-gain gating	Trait/state gain (NE)	Increase sensory noise; NE modulation; DCM of tDCS–EEG/fMRI	Friston (2005); Feldman & Friston (2010); Bastos et al. (2012)
Modulator	Homeostatic setpoint & hyperexcitability	High tonic gain brings network near nonlinear regime; small E↓ evokes large I↑; slow drift toward set-points	Participant-specific effects; session drift	Trait hyperexcitability; prior stimulation	Stratify by excitability; longitudinal tracking; WC + homeostasis fits	Hengen et al. (2013); Filmer et al. (2019); Godin et al. (2025)
Enabler	FEM-optimized multichannel tDCS (E targeting)	Right coverage/intensity reveals non-linearities; too focal/strong misses them	Effect size correlates with E norm/ERNI; optimized > 4×1	Anatomy; intensity; field orientation	In silico–in vivo correlation; 7-ch vs 4×1; polarity/intensity sweeps	Miranda et al. (2013); Ruffini et al. (2014); Ruffini et al. (2018); Joshi et al. (2024)

Having outlined these candidate frameworks, an important next step is to specify how they might be differentiated empirically (some examples are given in Table 2). A key prediction of the proposed framework is that modest suppression of excitatory drive should have minimal effects in low-gain states but produce disproportionately large inhibitory outcomes in high-gain states. More specifically, an ISN-like regime predicts that reducing excitatory drive can paradoxically increase inhibitory influence, particularly when the system is operating close to an unstable or high-gain regime. By contrast, gain-control accounts would predict a rescaling of responses through mechanisms such as divisive normalisation, whereby the response of a neural population is adjusted relative to the broader level of activity in the system. On this view, paradoxical suppression would reflect an attempt to re-centre activity within an optimal operating range, rather than a specific ISN-like shift in excitatory-inhibitory dynamics.

Predictive-coding accounts would make a different prediction: effects should scale primarily with uncertainty, reliability, or precision of the incoming signal, rather than with input amplitude alone. These alternatives could be tested by sorting participants according to baseline excitability and quantifying the relationship between perturbation strength, stimulus salience, uncertainty, and downstream autonomic suppression. This approach would enable a more direct assessment of whether paradoxical inhibition arises from local or distributed ISN-like circuit dynamics, broader gain-control and normalisation mechanisms, or precision-weighted inference, thereby evaluating the extent to which ISN-like principles can be generalised to macrocircuit function.

Implications for Clinical and Neuropsychiatric Interventions

The discussion fostered in the current paper also has implications for how null or inconsistent stimulation findings are interpreted in treatment and intervention-based work. A common response to weak effects in non-invasive brain stimulation research is to increase stimulation amplitude, extend stimulation duration, or add more repeated sessions, on the assumption that stronger or longer stimulation should produce larger and more reliable effects (Imperio & Chua, 2024; Jiang et al., 2024; Aust et al., 2022; Jones et al., 2018; Benninger et al., 2010). While such manipulations may be appropriate in some contexts, they can also reflect an overly linear view of neuromodulation. In this sense, failed stimulation effects should not automatically motivate "more stimulation"; they should motivate better-specified models of target engagement, network state, and expected direction of change.

If stimulation effects depend on cortical state, target anatomy, montage precision, baseline excitability, and network operating regime, then increasing dose may not simply amplify the intended effect (Esmaeilpour et al., 2018). It may shift the system into a different dynamical regime, recruit compensatory mechanisms, or obscure effects that only emerge within a specific range of perturbation (akin to a kind of 'sweet spot'). This is particularly important for studies that compare only anodal stimulation with sham, because such designs implicitly preserve the assumption that facilitation is the primary effect of interest and may miss cathodal, polarity-atypical, or nonlinear outcomes (see also Parkin et al., 2015 for related criticisms of unbalanced designs). A more theoretically grounded approach would treat polarity, intensity, timing, montage optimisation, and participant state as interacting variables rather than as independent levers for increasing effect size.

Conclusion

Atypical findings in brain stimulation research are not unprecedented. While often treated as anomalies, they may in fact point toward fundamental principles of cortical computation that remain poorly understood. Far from being mere curiosities, such effects challenge linear polarity-based assumptions and may signal new ways to conceptualise brain function and neuromodulation. In the present review, we have focused on a more specific form of paradoxical outcome: the possibility that suppressing excitatory drive within a regulatory / inhibitory cortical region can, under particular trait and state-dependent conditions, lead to stronger downstream autonomic inhibition rather than disinhibition.

We have outlined three theoretical frameworks to guide thinking in this area: precision-weighted predictive coding, gain control, and inhibition-stabilised networks (ISNs). Each framework can, in different ways, help explain how paradoxical inhibition might emerge when appropriately scaled to distributed neural systems. These models are complementary rather than mutually exclusive; they differ in biological specificity and explanatory scope but converge on a common theme: inhibitory responses are dynamically shaped by baseline state, contextual salience, network history, and the operating regime of the system.

Our prior work highlights cortical hyperexcitability as a key trait-level moderator of stimulation outcomes, a factor that may bias networks toward high-gain or unstable regimes where nonlinear inhibitory responses become more likely (Braithwaite et al., 2015; Joshi et al., 2024). Under such conditions, modest perturbations of excitatory drive may not simply reduce activity, but may alter the stability, reliability, or gain of signals propagating through cortico-limbic-autonomic circuits. This provides a plausible theoretical route by which

cathodal / inhibitory stimulation of the rVLPFC could produce greater suppression of SCRs, particularly during aversive / high-salience processing.

We do not claim that these findings establish paradoxical inhibition as a domain-general response to cortical suppression across stimulation protocols, cortical targets, or physiological outcomes. Rather, they identify a theoretically important boundary case that requires careful replication, computational formalisation, and targeted empirical testing. Future work will need to distinguish whether such effects are best explained by ISN-like circuit dynamics, broader gain-control and normalisation mechanisms, precision-weighted inference, or some interaction of these processes. In doing so, the field will need to move beyond polarity-based heuristics toward models grounded in cellular dynamics, distributed network interactions, inter-regional feedback, and the state-dependent operating conditions under which neuromodulation occurs.

Acknowledgements

The present work was supported by a research grant awarded to the primary author (JJB) by the UKRI / BBSRC funding body (APP61288). We thank the BBSRC for their generous support.

References

Abbott, L. F., & Chance, F. S. (2005). Drivers and modulators from push-pull and balanced synaptic input. *Progress in Brain Research*, *149*, 147-155. https://doi.org/10.1016/S0079-6123(05)49011-1

Abbott, L. F., Varela, J. A., Sen, K., & Nelson, S. B. (1997). Synaptic Depression and Cortical Gain Control. *Science*, *275*(5297), 221–224. https://doi.org/10.1126/science.275.5297.221

Anderson, J. S., Carandini, M., & Ferster, D. (2000). Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. *Journal of Neurophysiology*, *84*(2), 909–926. https://doi.org/10.1152/jn.2000.84.2.909.

Antal, A., Kincses, T. Z., Nitsche, M. A., Bartfai, O., & Paulus, W. (2004). Excitability changes induced in the human primary visual cortex by transcranial direct current stimulation: direct electrophysiological evidence. *Investigative Ophthalmology & Visual Science*, *45*(2), 702-707. https://doi.org/10.1167/iovs.03-0688

Arruda-Carvalho, M., & Clem, R. L. (2014). Pathway-selective adjustment of prefrontal-amygdala transmission during fear encoding. *Journal of Neuroscience*, *34*(47), 15601-15609. https://doi.org/10.1523/JNEUROSCI.2664-14.2014

Aston-Jones, G., & Cohen, J. D. (2005). An Integrative Theory of Locus Coeruleus-Noepinephrine Function: Adaptive Gain and Optimal Performance. *Annual Review of Neuroscience*, *28*, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Auksztulewicz, R., & Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex*, *80*, 125–140. https://doi.org/10.1016/j.cortex.2015.11.024

Aust, S., Brakemeier, E. L., Spies, J., Herrera-Melendez, A. L., Kaiser, T., Fallgatter, A., ... & Bajbouj, M. (2022). Efficacy of augmentation of cognitive behavioral therapy with transcranial direct current stimulation for depression: a randomized clinical trial. *JAMA Psychiatry*, *79*(6), 528-537. https://doi.org/ 10.1001/jamapsychiatry.2022.0696

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*(7), 419–429. https://doi.org/10.1038/nrn3950

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

Beissner, F., Meissner, K., Bär, K. J., & Napadow, V. (2013). The autonomic brain: an activation likelihood estimation meta-analysis for central processing of autonomic function. *Journal of Neuroscience*, *33*(25), 10503-10511. https://doi.org/10.1523/JNEUROSCI.1103-13.2013

Bell, S. B., Turner, B., Sawaki, L., & DeWall, N. (2022). When brain stimulation backfires: the effects of prefrontal cortex stimulation on impulsivity. *Social Cognitive and Affective Neuroscience*, *17*(1), 101-108. https://doi.org/10.1093/scan/nsaa049

Benninger, D. H., Lomarev, M., Lopez, G., Wassermann, E. M., Li, X., Considine, E., & Hallett, M. (2010). Transcranial direct current stimulation for the treatment of Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *81*(10), 1105-1111. https://doi.org/10.1136/jnnp.2009.202556

Benwell, C. S., Learmonth, G., Miniussi, C., Harvey, M., & Thut, G. (2015). Non-linear effects of transcranial direct current stimulation as a function of individual baseline performance: Evidence from biparietal tDCS influence on lateralized attention bias. *Cortex*, *69*, 152-165. https://doi.org/10.1016/j.cortex.2015.05.007

Borckardt, J. J., Bikson, M., Frohman, H., Reeves, S. T., Datta, A., Bansal, V., ... & George, M. S. (2012). A pilot study of the tolerability and effects of high-definition transcranial direct current stimulation (HD-tDCS) on pain perception. *The Journal of Pain*, *13*(2), 112-120. https://doi.org/10.1016/j.jpain.2011.07.001

Boucsein, W. (2012). *Electrodermal activity* (2nd ed.) New York, NY: Springer. https://doi.org/10.1007/978-1-4614-1126-0.

Braithwaite, J. J., Broglia, E., Brincat, O., Stapley, L., Wilkins, A. J., & Takahashi, C. (2013). Signs of increased cortical hyperexcitability selectively associated with spontaneous anomalous bodily experiences in a nonclinical population. *Cognitive Neuropsychiatry*, *18*(6), 549–573. https://doi.org/10.1080/13546805.2013.768176

Braithwaite, J. J., Mevorach, C., & Takahashi, C. (2015). Stimulating the aberrant brain: Evidence for increased cortical hyperexcitability from a transcranial direct current stimulation (tDCS) study of individuals predisposed to anomalous perceptions. *Cortex*, *69*, 1–13. https://doi.org/10.1016/j.cortex.2015.03.023

Cabibel, V., Muthalib, M., Teo, W. P., & Perrey, S. (2018). High-definition transcranial direct-current stimulation of the right M1 further facilitates left M1 excitability during crossed facilitation. *Journal of Neurophysiology*, *119*(4), 1266-1272. https://doi.org/10.1152/jn.00861.2017

Capogna, M., & Pearce, R. A. (2011). GABAA,slow: Causes and consequences. *Trends in Neurosciences*, *34*(2), 101–112. https://doi.org/10.1016/j.tins.2010.10.005

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*(1), 51–62. https://doi.org/10.1038/nrn3136

Chance, F. S., Abbott, L. F., & Reyes, A. D. (2002). Gain modulation from background synaptic input. *Neuron*, *35*(4), 773–782. https://doi.org/10.1016/S0896-6273(02)00820-6Cheng, J., Sahni, T., Zhao, Z., Stolte, S. E., You, C., Woods, A. J., ... & Fang, R. (2026). Recent Advancements of Transcranial Direct Current Stimulation and Machine Learning: Methods, Challenges, and Opportunities. *Transactions on Artificial Intelligence*, *2*(1), 54. https://doi.org/10.53941/tai.2026.100005

Cheng, S., Qiu, X., Li, S., Mo, L., Xu, F., & Zhang, D. (2022). Different roles of the left and right ventrolateral prefrontal cortex in cognitive reappraisal: an online transcranial magnetic stimulation study. *Frontiers in Human Neuroscience*, *16*, 928077. https://doi.org/10.3389/fnhum.2022.928077

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews Neuroscience*, *3*(8), 655-666. https://doi.org/10.1038/nrn894

Craig, A. D. (2003). Interoception: the sense of the physiological condition of the body. *Current Opinion in Neurobiology*, *13*(4), 500-505. https://doi.org/10.1016/S0959-4388(03)00090-4

Craig, A. D. (2009). How do you feel — now? The anterior insula and human awareness. *Nature Reviews Neuroscience, 10*, 59–70. https://doi.org/10.1038/nrn2555

Critchley, H. D. (2002). Electrodermal responses: what happens in the brain. *The Neuroscientist*, *8*(2), 132-142. https://doi.org/10.1177/107385840200800209

Critchley, H. D. (2002). Electrodermal responses: what happens in the brain. *The Neuroscientist*, *8*(2), 132-142. https://doi.org/10.1177/107385840200800209

Critchley, H. D., Elliott, R., Mathias, C. J., & Dolan, R. J. (2000). Neural activity relating to generation and representation of galvanic skin conductance responses: a functional magnetic resonance imaging study. *Journal of Neuroscience*, *20*(8), 3033-3040. https://doi.org/10.1523/JNEUROSCI.20-08-03033.2000

Datta, A., Elwassif, M., Battaglia, F., & Bikson, M. (2008). Transcranial current stimulation focality using disc and ring electrode configurations: FEM analysis. *Journal of Neural Engineering*, *5*(2), 163. https://doi.org/10.1088/1741-2560/5/2/007

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. *In J. T. Cacioppo, L. G. Tassinary, & G. G. Bemston (Eds.)*, *Handbook of Psychophysiology* (pp. 159–181). New York, NY: Cambridge University Press.

DeLaRosa, B. L., Spence, J. S., Motes, M. A., To, W., Vanneste, S., Hart Jr, J., & Kraut, M. A. (2020). The influence of pre-supplementary motor area targeted high-definition transcranial direct current stimulation on inhibitory control. *Biorxiv*, 2020-10. https://doi.org/10.1101/2020.10.27.358242

Denève, S., & Machens, C. K. (2016). Efficient codes and balanced networks. *Nature Neuroscience*, *19*(3), 375–382. https://doi.org/10.1038/nn.4243

Douglas, R. J., & Martin, K. A. C. (2004). Neuronal circuits of the neocortex. *Annual Review of Neuroscience*, *27*, 419–451. https://doi.org/10.1146/annurev.neuro.27.070203.144152.

Edwards, D., Cortes, M., Datta, A., Minhas, P., Wassermann, E. M., & Bikson, M. (2013). Physiological and modeling evidence for focal transcranial electrical brain stimulation in humans: a basis for high-definition tDCS. *Neuroimage*, *74*, 266-275. https://doi.org/10.1016/j.neuroimage.2013.01.042

Esmaeilpour, Z., Marangolo, P., Hampstead, B. M., Bestmann, S., Galletta, E., Knotkova, H., & Bikson, M. (2018). Incomplete evidence that increasing current intensity of tDCS boosts outcomes. *Brain Stimulation*, *11*(2), 310-321. https://doi.org/10.1016/j.brs.2017.12.002

Etkin, A., Egner, T., & Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in Cognitive Sciences*, *15*(2), 85–93. https://doi.org/10.1016/j.tics.2010.11.004

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 215. https://doi.org/10.3389/fnhum.2010.00215

Ferguson, K. A., & Cardin, J. A. (2020). Mechanisms underlying gain modulation in the cortex. *Nature Reviews Neuroscience*, *21*(2), 80–92. https://doi.org/10.1038/s41583-019-0253-y

Ffytche, D. H., Howard, R. J., Brammer, M. J., David, A. S., Woodruff, P. W. R., & Williams, S. C. R. (1998). The anatomy of conscious vision: An fMRI study of visual hallucinations. *Nature Neuroscience*, *1*(8), 738–742. https://doi.org/10.1038/3679.

Ffytche, D. H. (2008). The hodology of hallucinations. *Cortex*, *44*(8), 1067-1083. https://doi.org/10.1016/j.cortex.2008.04.005

Filmer, H. L., Ehrhardt, S. E., Bollmann, S., Mattingley, J. B., & Dux, P. E. (2019). Accounting for individual differences in the response to tDCS with baseline levels of neurochemical excitability. *Cortex*, *115*, 324–334. https://doi.org/10.1016/j.cortex.2019.02.012

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2008). Hierarchical Models in the Brain. *PLOS Computational Biology*, *4*(11), e1000211. https://doi.org/10.1371/journal.pcbi.1000211

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B*, *364*(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Froemke, R. C. (2015). Plasticity of cortical excitatory-inhibitory balance. *Annual Review of Neuroscience*, *38*, 195–219. https://doi.org/10.1146/annurev-neuro-071714-034002

Garnett, E. O., Malyutina, S., Datta, A., & den Ouden, D. B. (2015). On the use of the terms anodal and cathodal in high-definition transcranial direct current stimulation: a technical note. Neuromodulation: *Technology at the Neural Interface*, *18*(8), 705-713. https://doi.org/10.1111/ner.12320

Godin, C., Krause, M. R., Vieira, P. G., Pack, C. C., & Thivierge, J.-P. (2025). Control of Inhibition-Stabilized Oscillations in Wilson-Cowan Networks with Homeostatic Plasticity. *Entropy*, *27*(2), 215. https://doi.org/10.3390/e27020215

Gorrino, I., Canessa, N., & Mattavelli, G. (2023). Testing the effect of high-definition transcranial direct current stimulation of the insular cortex to modulate decision-making and executive control. *Frontiers in Behavioral Neuroscience*, *17*, 1234837. https://doi.org/10.3389/fnbeh.2023.1234837

Haider, B., & McCormick, D. A. (2009). Rapid neocortical dynamics: cellular and network mechanisms. *Nature Reviews Neuroscience*, *10*(7), 419–431. https://doi.org/10.1016/j.neuron.2009.04.008

Haider, B., Duque, A., Hasenstaub, A. R., & McCormick, D. A. (2006). Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *Journal of Neuroscience*, *26*(17), 4535–4545. https://doi.org/10.1523/JNEUROSCI.5297-05.2006

Harris, K. D., & Thiele, A. (2011). Cortical state and attention. *Nature Reviews Neuroscience*, *12*(9), 509–523. https://doi.org/10.1038/nrn3084

Hengen, K. B., Lambo, M. E., Van Hooser, S. D., Katz, D. B., & Turrigiano, G. G. (2013). Firing rate homeostasis in visual cortex of freely behaving rodents. *Neuron*, *80*(2), 335–342. https://doi.org/10.1016/j.neuron.2013.08.038.

Hsu, T. Y., Juan, C. H., & Tseng, P. (2016). Individual differences and state-dependent responses in transcranial direct current stimulation. *Frontiers in Human Neuroscience*, *10*, 643. https://doi.org/10.3389/fnhum.2016.00643

Hu, H., Gan, J., & Jonas, P. (2014). Fast-spiking, parvalbumin+ GABAergic interneurons: From cellular design to microcircuit function. *Science*, *345*(6196), 1255263. https://doi.org/10.1126/science.1255263

Imperio, C. M., & Chua, E. F. (2024). Lack of effects of online HD-tDCS over the left or right DLPFC in an associative memory and metamemory monitoring task. *PLoS One*, *19*(6), e0300779. https://doi.org/10.1371/journal.pone.0300779

Isaacson, J. S., & Scanziani, M. (2011). How inhibition shapes cortical activity. *Neuron*, *72*(2), 231–243. https://doi.org/10.1016/j.neuron.2011.09.027

Iwabuchi, S. J., Raschke, F., Auer, D. P., Liddle, P. F., Lankappa, S. T., & Palaniyappan, L. (2017). Targeted transcranial theta-burst stimulation alters fronto-insular network and prefrontal GABA. *Neuroimage*, *146*, 395-403. https://doi.org/10.1016/j.neuroimage.2016.09.043

Jacobson, L., Ezra, A., Berger, U., & Lavidor, M. (2012). Modulating oscillatory brain activity correlates of behavioral inhibition using transcranial direct current stimulation. *Clinical Neurophysiology*, *123*(5), 979-984. https://doi.org/10.1016/j.clinph.2011.09.016

Jay, E. L., Nestler, S., Sierra, M., McClelland, J., Kekic, M., & David, A. S. (2016). Ventrolateral prefrontal cortex repetitive transcranial magnetic stimulation in the treatment of depersonalization disorder: A consecutive case series. *Psychiatry Research*, *240*, 118-122. https://doi.org/10.1016/j.psychres.2016.04.027

Jay, E. L., Sierra, M., Van den Eynde, F., Rothwell, J. C., & David, A. S. (2014). Testing a neurobiological model of depersonalization disorder using repetitive transcranial magnetic stimulation. *Brain Stimulation*, *7*(2), 252-259. https://doi.org/10.1016/j.brs.2013.12.002

Jiang, S., Jones, M., & von Bastian, C. C. (2024). TDCS over PPC or DLPFC does not improve visual working memory capacity. *Communications Psychology*, 2(1), 20. https://doi.org/10.1038/s44271-024-00067-8

Joshi, S. D., Ruffini, G., Nuttall, H. E., Watson, D. G., & Braithwaite, J. J. (2024). Optimised Multi-Channel Transcranial Direct Current Stimulation (MtDCS) Reveals Differential Involvement of the Right-Ventrolateral Prefrontal Cortex (rVLPFC) and Insular Complex in those Predisposed to Aberrant Experiences. *Consciousness and Cognition*, *117*, 103610. https://doi.org/10.1016/j.concog.2023.103610

Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: Predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1668), 20140169. https://doi.org/10.1098/rstb.2014.0169

Kawaguchi, Y., & Kubota, Y. (1997). GABAergic cell subtypes and their synaptic connections in rat frontal cortex. *Cerebral Cortex*, *7*(6), 476–486. https://doi.org/10.1093/cercor/7.6.476

Kuo, H. I., Bikson, M., Datta, A., Minhas, P., Paulus, W., Kuo, M. F., & Nitsche, M. A. (2013). Comparing cortical plasticity induced by conventional and high-definition 4× 1 ring tDCS: a neurophysiological study. *Brain Stimulation*, *6*(4), 644-648. https://doi.org/10.1016/j.brs.2012.09.010

Li, L. M., Violante, I. R., Leech, R., Ross, E., Hampshire, A., Opitz, A., ... & Sharp, D. J. (2019). Brain state and polarity dependent modulation of brain networks by transcranial direct current stimulation. *Human Brain Mapping*, *40*(3), 904-915. https://doi.org/10.1002/hbm.24420

Li, Y., Ward, M. J., Richardson, R. M., G'Sell, M., & Ghuman, A. S. (2020). Endogenous activity modulates stimulus and circuit-specific neural tuning and predicts perceptual behavior. *Nature Communications*, *11*(1), 4014. https://doi.org/10.1038/s41467-020-17729-w

Litwin-Kumar, A., & Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature Neuroscience*, *15*(11), 1498–1505. https://doi.org/10.1038/nn.3220

Maldonado T, Goen JRM, Imburgio MJ, Eakin SM, Bernard JA (2019) Single session high definition transcranial direct current stimulation to the cerebellum does not impact higher cognitive function. *PLoS ONE*, *14*(10): e0222995. https://doi.org/10.1371/journal.pone.0222995

Medford, N. (2012). Emotion and the unreal self: depersonalization disorder and deaffectualization. *Emotion Review*, *4*(2), 139-144. https://doi.org/10.1177/1754073911430135

Millidge, B., Seth, A., & Buckley, C. L. (2022). Predictive Coding: A Theoretical and Experimental Review (arXiv:2107.12979). *arXiv*. https://doi.org/10.48550/arXiv.2107.12979

Millman, L. M., Huang, X., Wainipitapong, S., Medford, N., & Pick, S. (2024). Behavioural, autonomic, and neural responsivity in depersonalisation-derealisation disorder: A systematic review of experimental evidence. *Neuroscience & Biobehavioral Reviews*, *163*, 105783. https://doi.org/10.1016/j.neubiorev.2024.105783

Miranda, P. C., Callejón-Leblic, M. A., Salvador, R., & Ruffini, G. (2018). Realistic modeling of transcranial current stimulation: The electric field in the brain. *Current Opinion in Biomedical Engineering*, *8*, 20-27. https://doi.org/10.1016/j.cobme.2018.09.002

Miranda, P. C., Mekonnen, A., Salvador, R., & Ruffini, G. (2013). The electric field in the cortex during transcranial current stimulation. *NeuroImage, 70*, 48–58. https://doi.org/10.1016/j.neuroimage.2012.12.034

Muir, J., & Kittler, J. T. (2018). Plasticity of GABAA receptor diffusion dynamics at the axon initial segment. *Frontiers in Cellular Neuroscience*, *12*, 142. https://doi.org/10.3389/fncel.2014.00151

Nagai, Y., Critchley, H. D., Featherstone, E., Trimble, M. R., & Dolan, R. J. (2004). Activity in ventromedial prefrontal cortex covaries with sympathetic skin conductance level: a physiological account of a "default mode" of brain function. *Neuroimage*, *22*(1), 243-251. https://doi.org/10.1016/j.neuroimage.2004.01.019

Ozeki, H., Finn, I. M., Schaffer, E. S., Miller, K. D., & Ferster, D. (2009). Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron*, *62*(4), 578–592. https://doi.org/10.1016/j.neuron.2009.03.028

Parkin, B. L., Ekhtiari, H., & Walsh, V. F. (2015). Non-invasive human brain stimulation in cognitive neuroscience: a primer. *Neuron*, *87*(5), 932-945. https://doi.org/10.1016/j.neuron.2015.07.032

Phillips, M. L., Drevets, W. C., Rauch, S. L., & Lane, R. (2003). Neurobiology of emotion perception I: The neural basis of normal emotion perception. *Biological Psychiatry*, *54*(5), 504-514. https://doi.org/10.1016/S0006-3223(03)00168-9

Phillips, M. L., Medford, N., Senior, C., Bullmore, E. T., Suckling, J., Brammer, M. J., ... & David, A. S. (2001). Depersonalization disorder: thinking without feeling. *Psychiatry Research: Neuroimaging*, *108*(3), 145-160. https://doi.org/10.1016/S0925-4927(01)00119-6

Rabinowitz, N. C., Goris, R. L., Cohen, M., & Simoncelli, E. P. (2015). Attention stabilizes the shared gain of V4 populations. *eLife*, *4*, e08998. https://doi.org/10.7554/eLife.08998

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Rubenstein, J. L., & Merzenich, M. M. (2003). Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes, Brain and Behavior*, *2*(5), 255–267. https://doi.org/10.1034/j.1601-183X.2003.00037.x

Rubin, D. B., Van Hooser, S. D., & Miller, K. D. (2015). The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron*, *85*(2), 402–417. https://doi.org/10.1016/j.neuron.2014.12.026

Ruffini, G., Salvador, R., Castaldo, F., Baleeiro, T., Camprodon, J. A., Chopra, M., Cappon, D. and Pascual-Leone, A. (2024). Multichannel tDCS with advanced targeting for major depressive disorder: a tele-supervised at-home pilot study. *Frontiers in Psychiatry 15*,1427365. https://doi.org/10.3389/fpsyt.2024.1427365

Ruffini, G., Fox, M. D., Ripolles, O., Miranda, P. C., & Pascual-Leone, A. (2014). Optimization of multifocal transcranial current stimulation for weighted cortical pattern targeting from realistic modeling of electric fields. *NeuroImage, 89*, 216–225. https://doi.org/10.1016/j.neuroimage.2013.12.002

Ruffini, G., Wendling, F., Sanchez-Todo, R., & Santarnecchi, E. (2018). Targeting brain networks with multichannel transcranial current stimulation (tCS). *Current Opinion in Biomedical Engineering, 8*, 70–77. https://doi.org/10.1016/j.cobme.2018.11.001

Salehinejad, M. A., Abdi, M., Dadashi, M., Habibi, M., Hallajian, A. H., Sharifi, K., ... & Nitsche, M. A. (2026). Optimized multichannel 4 mA vs conventional transcranial direct current stimulation for major depressive disorder: A randomized sham-controlled trial. *Molecular Psychiatry*, 1-13. https://doi.org/10.1038/s41380-026-03560-0

Salvador, R., Mekonnen, A., Ruffini, G., & Miranda, P. C. (2010, August). Modeling the electric field induced in a high resolution realistic head model during transcranial current stimulation. *In 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology* (pp. 2073-2076). IEEE. https://doi.org/10.1109/IEMBS.2010.5626315

Sanzeni, A., Akitake, B., Goldbach, H. C., Leedy, C. E., Brunel, N., & Histed, M. H. (2020). Inhibition stabilization is a widespread property of cortical networks. *eLife*, *9*, e54875. https://doi.org/10.7554/eLife.54875

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience*, *4*(8), 819–825. https://doi.org/10.1038/90526

Servan-Schreiber, D., Printz, H., & Cohen, J. D. (1990). A Network Model of Catecholamine Effects: Gain, Signal-to-Noise Ratio, and Behavior. *Science*, *249*(4971), 892–895. https://doi.org/10.1126/science.2392679

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–573. https://doi.org/10.1016/j.tics.2013.09.007

Shadlen, M. N., & Newsome, W. T. (1994). Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, *4*(4), 569–579. https://doi.org/10.1016/0959-4388(94)90059-0.

Sierra, M., & Berrios, G. E. (1998). Depersonalization: neurobiological perspectives. *Biological Psychiatry*, *44*(9), 898-908. https://doi.org/10.1016/S0006-3223(98)00015-8

Sierra, M., & David, A. S. (2011). Depersonalization: a selective impairment of self-awareness. *Consciousness and Cognition*, *20*(1), 99-108. https://doi.org/10.1016/j.concog.2010.10.018

Sierra, M., Lopera, F., Lambert, M. V., Phillips, M. L., & David, A. S. (2002). Separating depersonalisation and derealisation: the relevance of the "lesion method". *Journal of Neurology, Neurosurgery & Psychiatry*, *72*(4), 530-532. http://doi.org/10.1136/jnnp.72.4.530

Sierra-Mercado, D., Padilla-Coreano, N., & Quirk, G. J. (2011). Dissociable roles of prelimbic and infralimbic cortices, ventral hippocampus, and basolateral amygdala in the expression and extinction of conditioned fear. *Neuropsychopharmacology*, *36*(2), 529-538. https://doi.org/10.1038/npp.2010.184

Silberberg, G., & Markram, H. (2007). Disynaptic Inhibition between Neocortical Pyramidal Cells Mediated by Martinotti Cells. *Neuron*, *53*(5), 735–746. https://doi.org/10.1016/j.neuron.2007.02.012

Soldado-Magraner, S., Seay, M. J., Laje, R., & Buonomano, D. V. (2022). Paradoxical self-sustained dynamics emerge from orchestrated excitatory and inhibitory homeostatic plasticity rules. *Proceedings of the National Academy of Sciences*, *119*(43), e2200621119. https://doi.org/10.1073/pnas.2200621119

Spratling, M. W. (2008). Reconciling predictive coding and biased competition models of cortical function. *Frontiers in Computational Neuroscience*, *2*. https://doi.org/10.3389/neuro.10.004.2008

Spratling, M. W. (2010). Predictive Coding as a Model of Response Properties in Cortical Area V1. *Journal of Neuroscience*, *30*(9), 3531–3543. https://doi.org/10.1523/JNEUROSCI.4911-09.2010

Tamás, G., Lőrincz, A., Simon, A., & Szabadics, J. (2003). Identified Sources and Targets of Slow Inhibition in the Neocortex. *Science*, *299*(5614), 1902–1905. https://doi.org/10.1126/science.1082053

Tremblay, R., Lee, S., & Rudy, B. (2016). GABAergic interneurons in the neocortex: from cellular properties to circuits. *Neuron*, *91*(2), 260–292. https://doi.org/10.1016/j.neuron.2016.06.033

Tsodyks, M. V., Skaggs, W. E., Sejnowski, T. J., & McNaughton, B. L. (1997). Paradoxical effects of external modulation of inhibitory interneurons. *The Journal of Neuroscience*, *17*(11), 4382–4388. https://doi.org/10.1523/JNEUROSCI.17-11-04382.1997

Urban-Ciecko, J., & Barth, A. L. (2016). Somatostatin-expressing neurons in cortical networks. *Nature Reviews Neuroscience*, *17*(7), 401–409. https://doi.org/10.1038/nrn.2016.53

Vergallito, A., Varoli, E., Pisoni, A., Mattavelli, G., Del Mauro, L., Feroldi, S., Vallar, G., & Romero Lauro, L. J. (2023). State-dependent effectiveness of cathodal transcranial direct current stimulation on cortical excitability. *NeuroImage*, *277*, 120242. https://doi.org/10.1016/j.neuroimage.2023.120242

Villamar, M. F., Wivatvongvana, P., Patumanond, J., Bikson, M., Truong, D. Q., Datta, A., & Fregni, F. (2013). Focal modulation of the primary motor cortex in fibromyalgia using 4× 1-ring high-definition transcranial direct current stimulation (HD-tDCS): immediate and delayed analgesic effects of cathodal and anodal stimulation. *The Journal of Pain*, *14*(4), 371-383. https://doi.org/10.1016/j.jpain.2012.12.007

Vogels, T. P., & Abbott, L. F. (2009). Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. *Nature Neuroscience*, *12*(4), 483–491. https://doi.org/10.1038/nn.2276

Wilson, N. R., Runyan, C. A., Wang, F. L., & Sur, M. (2012). Division and subtraction by distinct cortical inhibitory networks in vivo. *Nature*, *488*(7411), 343–348. https://doi.org/10.1038/nature11347

Wu, Y. K., & Gjorgjieva, J. (2023). Inhibition stabilization and paradoxical effects in recurrent neural networks with short-term plasticity. *Physical Review Research*, *5*(3), 033023. https://doi.org/10.1103/PhysRevResearch.5.033023

Wyss, A. M., Baumgartner, T., Guizar Rosales, E., Soutschek, A., & Knoch, D. (2024). Cathodal HD-tDCS above the left dorsolateral prefrontal cortex increases environmentally sustainable decision-making. *Frontiers in Human Neuroscience*, *18*, 1395426. https://doi.org/10.3389/fnhum.2024.1395426

Yizhar, O., Fenno, L. E., Prigge, M., Schneider, F., Davidson, T. J., O'shea, D. J., Sohal, V. S., Goshen, I., Finklestein, J., Paz, J. T., Stehfest, K., Fudim, R., Ramakrishnan, C., Huguenard, J. R., Hegemann, P. & Deisseroth, K. (2011). Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature*, *477*(7363), 171-178. https://doi.org/10.1038/nature10360

Zhang, J., & Abbott, L. F. (2000). Gain modulation of recurrent networks. *Neurocomputing*, *32–33*, 623–628. https://doi.org/10.1016/S0925-2312(00)00224-1