Causal mapping of human brain function

Shan H. Siddiqi [ID][1,2✉], *Konrad P. Kording[3], Josef Parvizi* [ID][4] *and Michael D. Fox[1,5]*

Abstract | Mapping human brain function is a long-standing goal of neuroscience that promises to inform the development of new treatments for brain disorders. Early maps of human brain function were based on locations of brain damage or brain stimulation that caused a functional change. Over time, this approach was largely replaced by technologies such as functional neuroimaging, which identify brain regions in which activity is correlated with behaviours or symptoms. Despite their advantages, these technologies reveal correlations, not causation. This creates challenges for interpreting the data generated from these tools and using them to develop treatments for brain disorders. A return to causal mapping of human brain function based on brain lesions and brain stimulation is underway. New approaches can combine these causal sources of information with modern neuroimaging and electrophysiology techniques to gain new insights into the functions of specific brain areas. In this Review, we provide a definition of causality for translational research, propose a continuum along which to assess the relative strength of causal information from human brain mapping studies and discuss recent advances in causal brain mapping and their relevance for developing treatments.

Causality gap
The missing logical link between correlation and causation, particularly as it relates to identifying treatment targets.

[1]Center for Brain Circuit Therapeutics, Brigham & Women's Hospital, Boston, MA, USA.

[2]Department of Psychiatry, Harvard Medical School, Boston, MA, USA.

[3]Department of Neuroscience, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA.

[4]Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Palo Alto, CA, USA.

[5]Department of Neurology, Harvard Medical School, Boston, MA, USA.

✉e-mail: shsiddiqi@bwh.harvard.edu

https://doi.org/10.1038/s41583-022-00583-8

Human brain mapping studies often rely on neurophysiological correlates of symptoms or behaviour[1–4]. This approach has yielded a wealth of knowledge about brain organization[5–8], but this knowledge has not routinely translated into therapeutic targets[3,9–11]. The difference between correlation and causation may be an important reason for this lack of translation[12,13]. Correlative methods do not distinguish between brain regions that are causing a symptom, compensating for a symptom or incidentally related to a symptom. As such, interventions based on neurophysiological correlates could improve a symptom, worsen a symptom or have no effect.

This 'causality gap' is a well-recognized problem in human neuroscience[10,13,14], and echoes a problem that other disciplines have also faced. For instance, the field of econometrics experienced a 'credibility revolution'[15] after an influential push to use causal methods in the early 1980s (REF.[16]). By emulating this example, the causality gap in neuroscience may be better framed as a causality opportunity. Wider adoption of causal methods could improve our ability to translate brain mapping findings into successful interventions[13,14,17,18]. Causal mapping studies using brain lesions and brain stimulation helped lead to mainstream clinical treatments such as deep brain stimulation (DBS) for Parkinson disease[19–22] and transcranial magnetic stimulation (TMS) for major depression[23,24]. These successes may provide a road map

for identifying new therapeutic targets based on causal mapping of human brain function.

In this Review, we provide a clinically relevant definition of causality, summarize different approaches to causal brain mapping, propose a framework for appraising the strength of causality in brain mapping studies and describe how these studies can be translated into anatomically targeted treatments for neuropsychiatric disorders.

Brain mapping and causal inference

A brief history of human brain mapping. Early brain maps were based on lesions or stimulation sites that directly caused a functional change, ranging from influential lesion cases such as Phineas Gage[25] and H.M.[26] to systematic experiments such as Wilder Penfield's intracranial stimulation studies[27–29]. However, in the latter part of the twentieth century, the advent of functional neuroimaging[1,2,4,30] was accompanied by a decline in lesion and stimulation studies[31] (FIG. 1). Neuroimaging allowed non-invasive mapping of human brain function in patients without lesions, including healthy individuals. Neuroimaging correlates often aligned with prior lesion and stimulation results[32], and provided additional insights into complex functions such as language and memory[33,34].

Nevertheless, this new capability came with a cost, as these correlative tools moved away from the causal inference that was possible with brain lesions and
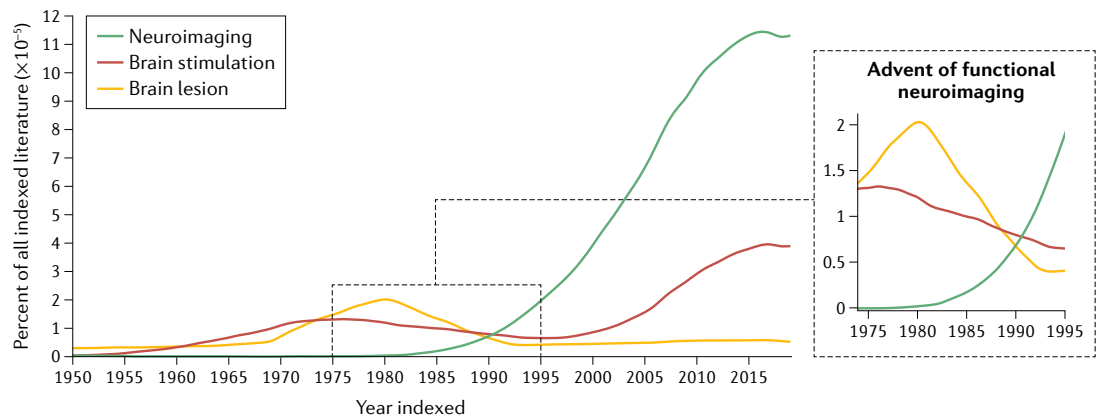
Fig. 1 | **Evolution of brain mapping literature since 1950.** In the late twentieth century, growth of neuroimaging literature coincided with a decline in brain lesion and brain stimulation literature. Plotted data obtained by conducting a Google Ngram search using the terms 'neuroimaging', 'brain stimulation' and 'brain lesion'[31], with data binned to show 1950–2019 and smoothing set to 5.

Deep brain stimulation
(DBS). An invasive technique for modulating brain activity using surgically implanted depth electrodes, commonly used as a treatment for Parkinson disease.

Transcranial magnetic stimulation
(TMS). A non-invasive technique for modulating brain activity via magnetic pulses applied over the scalp, commonly used as a treatment for major depressive disorder.

Functional neuroimaging
A non-invasive technique for measuring changes in brain activity, commonly based on changes in blood oxygenation or metabolic activity.

Causal inference
The process of assessing whether one event directly contributed to the occurrence of another event.

Counterfactual
The hypothetical outcome if a causal event had not occurred.

Causal chain
A sequence of events leading to a particular outcome.

Bradford Hill criteria
A widely used framework for objectively appraising the strength of causal inference based on observed data.

Randomized controlled trials
Research studies that randomly allocate clinical participants to active or control interventions.

---

brain stimulation[9,14,17,18,30]. Some saw this as a minor limitation[35], but others criticized functional neuroimaging as 'the new phrenology'[36]. More importantly, despite its popularity, functional neuroimaging was not leading to the development of new clinical treatments[14,17,18]. This led to essays highlighting the causality gap in human neuroscience[14,17] and the need to combine functional neuroimaging with causal techniques[17,18].

A brief history of causal inference. The history of causal inference has largely focused on addressing one fundamental problem — to conclude that an intervention caused an outcome, one must determine what would have happened if the intervention had been absent[37,38]. The counterfactual, or what would happen if an activity had been different, is unobservable by definition. Therefore, definitive causal inference is impossible. Instead, researchers make careful assumptions to estimate the counterfactual by studying the effects of a perturbation, such as a focal disruption of brain activity[10,39].

The value of understanding causality was recognized by philosophers as early as Aristotle, who believed that true knowledge is unattainable without asking "why"[40,41]. Aristotle's tradition inspired medieval theologians such as the Islamic physician–philosopher ibn Sina and the Catholic priest Thomas Aquinas[42]. Both ibn Sina and Aquinas reasoned that any causal chain either can be traced infinitely or should end in a primary cause, which Aquinas labelled as 'God'[42]. To explore the concept of free will, ibn Sina also described secondary causes that occur further down the chain set in place by God. As a physician, ibn Sina emphasized the need to modify outcomes by manipulating these secondary causes[43].

Philosophers debated this topic for centuries until David Hume codified eight logical 'rules' for causal inference in 1748 (REF.[44]). This early attempt inspired subsequent systems that were concrete enough to apply scientifically. For instance, Robert Koch's microbiological 'postulates' were successfully used to identify the aetiology of tuberculosis and cholera[45]. More recently, Austin Bradford Hill's epidemiological 'criteria' revealed many important causal relationships, such as the link between smoking and lung cancer[46–48].

The Bradford Hill criteria include temporality (the only mandatory criterion), specificity, effect size, reproducibility, dose–response, physiological plausibility, experimental manipulation, analogy, reversibility (an optional criterion) and coherence at different levels. The link between smoking and lung cancer exemplifies how these criteria can be used to establish a causal relationship: smoking consistently precedes lung cancer (temporality); lung cancer increases specifically after smoking, but not after related exposures such as chewing tobacco and alcohol (specificity); smoking has a large effect on the risk of lung cancer (effect size); the relationship between smoking and lung cancer is consistent across different settings (reproducibility); more smoking leads to greater lung cancer risk (dose–response); smoking causes single-strand DNA breaks, which predispose to neoplasia (physiological plausibility); in animal models, smoking increases cancer risk more than control interventions (experimental manipulation); the effects of smoking are similar to those of other known carcinogens (analogy); some smoking-induced changes are reversible upon cessation of smoking (reversibility); and these criteria are consistently satisfied by in vitro studies, animal studies, human studies and epidemiological studies (coherence).

Causal inference in biomedical science. The medical sciences did not explicitly focus on causality until ibn Sina's vision of controlled clinical trials[49] was finally realized in the mid-twentieth century[50]. Randomized trials changed the face of clinical therapeutics because they established causality by minimizing confounders, as the control group is identical to the intervention group in nearly every way except for the intervention itself. Randomized controlled trials thus became the gold standard for estimating the counterfactual by modelling a situation in which the intervention was absent[50].

However, before designing a clinical trial, one must identify components of the causal chain as potential treatment targets[51] — otherwise, there is a risk of intervening on the basis of misleading associations[52].

For instance, tuberculosis rates correlate with malaria rates, body temperature, mycobacterial infection and travel to endemic countries[53]. This might lead to the incorrect conclusion that tuberculosis could be treated with antimalarials or antipyretics. In reality, the correlation with malaria is spurious, as both diseases are endemic to lower-income countries, and the correlation with body temperature owes itself to reverse causation, as tuberculosis causes fever. By contrast, mycobacterial infection is causal, as the counterfactual has been demonstrated with confidence using Koch's postulates[54]. Travel to endemic countries is also in the causal chain of events, but only because it increases the risk of mycobacterial infection. Thus, before the first randomized clinical trial evaluated antimycobacterial therapy for tuberculosis, the likely outcome was already well understood[50].

When experimental manipulation in a randomized controlled trial is impractical or unethical, causality can still be assessed observationally using systematic approaches such as the Bradford Hill criteria[55]. Importantly, these observational approaches do not necessarily rule out all unmeasured confounders. For instance, it is plausible that the link between smoking and lung cancer is explained by an unobserved genetic factor that independently increases the risk for both. Econometricians often overcome this limitation by identifying natural experiments, in which an intervention is applied in a manner that is practically random. For instance, if lung cancer rates decline consistently in different districts that raise tobacco taxes, this effect is unlikely to be mediated by an unobserved genetic factor. Accordingly, natural experiments can be used to infer causality[56,57].

Causal inference in human brain mapping. Human brain mapping researchers often attempt causal inference to achieve one of two objectives. This Review focuses on the objective of symptom localization, which aims to identify causal links between symptoms and neuroanatomy[58]. When a symptom is successfully localized, it may potentially be treated by modulating the corresponding neuroanatomy. The second objective, mapping the direction of information flow, aims to understand how one brain region can causally influence another. These experiments attempt to estimate the direction of information flow between two or more nodes in the brain using various measures of effective connectivity[10]. Different methods for symptom localization are summarized in BOX 1, whereas different methods for measuring effective connectivity are summarized in BOX 2.

For any of these methods to provide causal knowledge, the study must be designed to estimate what would have happened if the intervention had been different[40]. When direct experimental manipulation is not possible, this may still be accomplished by regressing out potential confounders, fitting progressively larger models that contain the interactions or estimating how different factors influence the probability of an outcome[59]. These approaches differ in how they estimate the counterfactual, but can be combined into a single axiomatic definition for causality: an event is causal if, with all else being equal, its presence or absence affects the

probability of an outcome[10,37,60]. Even if the cause is not necessary or sufficient to induce the effect, this probabilistic definition can still reveal treatment targets. For instance, smoking does not always cause lung cancer and lung cancer is not always caused by smoking, but smoking single-handedly increases the risk of lung cancer. Following this definition, we can reduce the risk of lung cancer with interventions that reduce smoking.

Although the fundamental goal of this approach is to estimate the counterfactual, this is sometimes not enough to clearly demonstrate causality. For instance, task-based functional MRI (fMRI) experiments often compare the effects of a task with control conditions that estimate what would have happened if the task was not conducted. However, for the reasons described in BOX 1, these experiments are rarely used to argue for causality.

In addition to estimating the counterfactual, causal inference in brain mapping may be assessed by borrowing criteria from Bradford Hill's framework[61,62] (FIG. 2). Of these criteria, six are particularly useful for assessing the strength of causal inference across different human brain mapping studies: specificity, experimental manipulation, dose–response, coherence/convergence, reversibility and temporality. Other criteria, including effect size, reproducibility, plausibility and analogy, can be studied using any method, and thus are less useful for comparing the strength of causal inference across brain mapping studies.

When such criteria are applied rigorously, causal inference from observational data can approach the strength of causal inference from randomized controlled trials[57]. However, there is no established system for estimating the strength of causal inference from human brain mapping data. Building on Hume's framework and the many systematic approaches that it inspired, we suggest a 'causality continuum' for human brain mapping (FIG. 2b). Although it is impossible to conclusively determine the causal role of an intervention, this continuum may be applied to score the relative strength of causality for different brain mapping techniques.

Conceptual causal mapping frameworks

Among different human brain mapping approaches, brain lesions and brain stimulation are the furthest along the causality continuum, assuming that a study using one of these approaches is rigorously designed to estimate the counterfactual and to demonstrate specificity[10,18]. However, the number of possible lesions or stimulation sites is essentially infinite, as is the number of possible behavioural outcomes, making it impractical to prospectively disentangle the behavioural effects of every single brain region from all other brain regions[30]. This practical challenge can be overcome by identifying natural experiments with incidental or near-random variance in the location of the lesion or the stimulation site. Using large datasets, it is possible to test for anatomical specificity by comparing hundreds of incidental lesions or stimulation sites across the brain[63]. This can yield a large-scale map of brain regions associated with certain lesion-induced symptoms[64,65] or stimulation-induced responses[66]. This is commonly achieved by comparing the locations of lesion or stimulation locations that modify a symptom,

Natural experiments
Observed situations in which an intervention was incidentally assigned in a near-random manner.

Symptom localization
The process by which a neurological or psychiatric problem is linked to a specific brain region.

Effective connectivity
The influence that one neural unit exerts over another.

Box 1 | **Neuroimaging techniques for mapping symptoms or behaviours**

Some symptom localization approaches can yield stronger causal inference than others (FIG. 2). Of note, many correlative techniques can provide useful complementary information when combined with causal techniques[173].

Task-free neuroimaging

Task-free neuroimaging measures the brain at rest. Blood flow can be measured using positron emission tomography[2] or arterial spin labelling[174], connectivity can be estimated using resting-state functional MRI (fMRI) and diffusion tensor imaging[2], structure can be measured using volumetric MRI[57] and electrical activity can be measured using electroenceph-alography or magnetoencephalography[175]. These measurements can be compared with clinical outcomes to identify neuroanatomical correlates of a behaviour. Although task-free neuroimaging does not satisfy any of the criteria for causal inference[29,45], it can be combined with causal techniques to better understand the circuit-based mechanisms of a causal relationship[55]. Task-free neuroimaging can also be used to reliably map brain organization, potentially revealing targets for causal manipulation[173].

Task-based neuroimaging

Task-based neuroimaging measures the functional neuroanatomical correlates of a behavioural task. Studies using this approach estimate the counterfactual by comparing an experimental task with carefully designed control tasks. However, task-based imaging studies are rarely used to argue for causality owing to the temporal direction of the experiment — rather than manipulating brain activity and observing the resulting behaviour, these studies manipulate the behaviour and observe the resulting brain activity[3].

Neurofeedback

Neurofeedback is a variant of task-based neuroimaging in which a participant consciously manipulates their own task-based brain activity. As such, this technique modifies brain activity and measures a change in behaviour, which allows for stronger causal inference than traditional task-based neuroimaging, which modifies behaviour and measures changes in brain activity[176,177].

Incidental atrophy

Incidental atrophy patterns can be compared with behavioural outcomes, potentially illustrating specificity (assuming that the atrophy occurs in a near-random pattern) and dose–response (if greater atrophy leads to greater symptom severity). However, temporality is difficult to assess because atrophy progresses slowly. It can be unclear whether the atrophy caused the symptom, the symptom caused the atrophy or both were caused by an unmeasured variable. For instance, substantia nigra atrophy is associated with anosmia because both phenomena can be caused by Parkinson disease[178].

Incidental lesions

Incidental lesions can illustrate specificity and dose–response in the same way as incidental atrophy. Furthermore, because lesions occur rapidly, their temporal effects are usually clear. Some lesions are even reversible, as in multiple sclerosis. Thus, lesion studies can provide reasonably strong causal inference[3,10,30,58,63]. However, because lesions occur in an uncontrolled setting, it is difficult to apply experimental manipulations or to estimate the effect of unmeasured pre-lesion factors[3,17,29]. Lesions can also be challenging to interpret because of diaschisis or downstream effects on distantly connected brain regions[146].

Incidental stimulation sites

Incidental stimulation sites exist because therapeutic brain stimulation is often applied imprecisely to approximate targets. For example, transcranial magnetic stimulation (TMS) is usually targeted using scalp landmarks, leading to incidental variability in the stimulation site on the brain. When the actual stimulation sites are localized retrospectively, near-random variability is revealed in the precise region that is stimulated in any given patient[84,95]. Thus, the counterfactual can be estimated by comparing the post-treatment state with the pretreatment state, whereas specificity can be demonstrated by comparing effective with ineffective stimulation sites. However, this approach still relies on incidental variance (a natural experiment) rather than prospective experimental manipulation, and thus may be subject to unmeasured confounders.

Targeted lesions

Targeted lesions are often created surgically as a therapeutic intervention. In addition to the insights gleaned from incidental lesions, prospectively targeted lesions can be experimentally manipulated, although this is rare given that the procedure is invasive and generally irreversible[179,180]. This enables strong causal inference, but is currently limited by the inability to assess reversibility.

Targeted stimulation

Targeted stimulation has many of the same advantages as targeted lesions, with the added benefit of reversibility. It also has the same advantages as incidental stimulation sites, with the added benefit of enabling precise experimental manipulation. Thus, targeted stimulation provides data about nearly all of the criteria on the causality continuum[10,14,17,18]. However, there are often assumptions about the effect of stimulation on neural function that may not be accurate (for example, excitation versus inhibition). Furthermore, to illustrate coherence, targeted stimulation must be complemented with other techniques.

Convergent causal mapping

Convergent causal mapping incorporates multiple causal approaches into a single analysis. For instance, if a symptom is caused by lesioning a specific circuit and the same symptom is relieved by stimulating the same circuit, this cross-modal coherence enables stronger causal inference than either modality alone.

Probabilistic causality
An approach that assesses whether an event modifies the probability of an outcome, even if it is not necessary or sufficient to induce the outcome.

Voxel lesion symptom mapping
(VLSM). A probabilistic technique that analyses the relationship between tissue damage and behaviour on a voxel-by-voxel basis.

Diaschisis
The phenomenon by which damage to one brain region can cause a change in function of distant regions.

by comparing the connectivity of these lesions or stimulation sites, or by modifying symptoms in real time with direct stimulation and recording.

Mapping near-random brain lesions. Most early lesion and stimulation studies were based on case reports or small case series in which a lesion or a stimulation was sufficient to cause a certain outcome. These studies provided useful knowledge about the neuroanatomy of clearly measurable functions, such as sensation, movement and vision. However, some higher-order functions are difficult to study using individual cases, as stimulation of higher association areas does not always induce a measurable response[67]. This limitation can be mitigated using a probabilistic approach to causality, in which it is determined how a lesion or a stimulation modifies the probability of the outcome[63]. Probabilistic causality still lends itself to identifying therapeutic targets, as a treatment that reduces the probability of a symptom can still be clinically useful[37].

In a probabilistic study, large datasets are used to compare the effects of multiple lesions or stimulation sites. This was first illustrated with voxel lesion symptom mapping (VLSM), which compares symptom-causing lesions with other lesions[63]. For example, in retired US military personnel who sustained penetrating brain injury, lesions to the amygdala and anterior temporal lobe reduced the risk of post-traumatic stress disorder[68], whereas lesions to the posterior temporal lobe increased the risk of anxiety[69]. The probabilistic approach used in VLSM has also been expanded to brain stimulation studies — for instance, in patients receiving prefrontal TMS for major depression, relatively anterior and lateral stimulation sites are more effective than relatively posterior and medial sites[70–72]. However, the statistical power of VLSM is limited because most lesions and stimulation sites do not overlap with one another[58] (FIG. 3a), and so large sample sizes are required to detect statistical associations at any given brain location. VLSM also does not directly account for the role of diaschisis and interconnected brain networks[30,58], although network-level relationships may become apparent with sufficient spatial coverage and sample size.

Localizing symptoms to brain circuits. When similar symptoms are induced by non-overlapping lesions in different patients, it can be challenging to localize the

---

Box 2 | Mapping causal interactions between brain regions

Measuring effective connectivity using brain stimulation
Rather than assessing how certain brain regions causally influence specific behaviours, some studies assess how certain brain regions causally influence other brain regions. Causality can be inferred by combining focal stimulation with distant real-time physiological recordings[173]. For instance, transcranial magnetic stimulation (TMS) or deep brain stimulation (DBS) can be paired with immediate functional MRI (fMRI) or electroencephalography. These combinations are high on the causality continuum, as they employ targeted stimulation.

These real-time experiments investigate effective connectivity, which measures the direction of information flow in a brain circuit. If the experiments are rigorously controlled, then alternative and sham stimulation conditions can be used to estimate the counterfactual[173]. For instance, applying excitatory TMS versus inhibitory TMS to the default mode network (DMN) leads to dissociable effects on DMN connectivity[136,143]. Similarly, applying excitatory TMS to the frontoparietal control/central executive network can induce immediate changes in connectivity to the DMN, whereas inhibiting the frontoparietal control/central executive network or stimulating other networks does not induce the same changes[137]. This capability is consistent across modalities, as intracranial electrical stimulation (iES) elicits distinct electrophysiological responses when stimulating the DMN versus the frontoparietal control/central executive network or the salience network[181]. This principle of using alternate stimulation targets to estimate the counterfactual also extends to other networks, as TMS to the dorsal attention network versus the DMN can induce network-specific electroencephalography changes[138]. Collectively, these findings also suggest that TMS may be able to indirectly modify deeper brain regions that cannot be accessed directly.

Effective connectivity can provide mechanistic insights into direct or indirect relationships between sites of interest, and possible directionality and/or timing of signal flow between them. However, it does not offer causal inference about the importance of any of the sites for specific behaviours or functional outcomes[182].

Indirectly estimating connectivity
Effective connectivity may be estimated indirectly based on temporal relationships between activity in different brain regions[10]. These temporal forecasting methods are particularly useful for data with high temporal precision, such as intracranial measurements of local field potentials at the neuronal level[183]. The most common temporal forecasting approaches are Granger causality and dynamic causal modelling. Granger causality, which was originally designed for economic time series analysis, quantifies the directional relationship between two variables based on a temporal delay in their activity[184]. Dynamic causal modelling uses a similar principle, but also incorporates multiple regions into a probabilistic graphical model allowing for different hypotheses to be specified by varying its nodes and directed edges. The downstream effects of an incidental fluctuation or behavioural task can then be estimated by selecting the simplest explanatory model[185].

These indirect approaches provide weaker causal inference than brain stimulation experiments, as temporal forecasting is not immune to spurious relationships[10,186,187]. Countless unmeasured confounders could cause a temporal delay[188], particularly for techniques with lower temporal resolution such as functional neuroimaging[10]. Nevertheless, Granger causality and dynamic causal modelling are often misunderstood as tools with strong causal inference, possibly owing to misleading nomenclature[189,190].

The relationship between brain regions can also be estimated using task-free neuroimaging techniques such as resting-state functional connectivity and diffusion tractography. These approaches are low on the causality continuum, and provide no information about the causal direction of information flow.
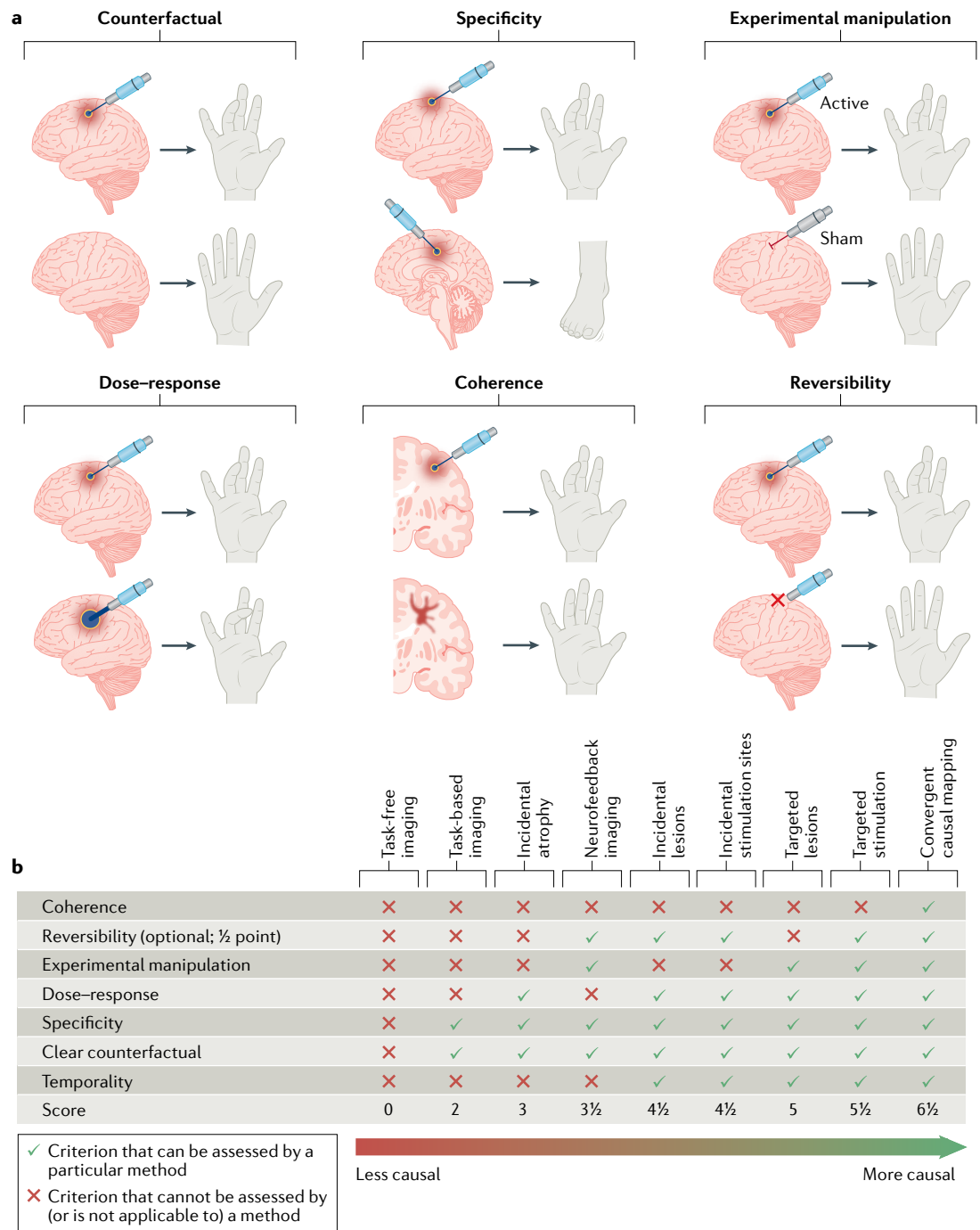
Fig. 2 | **Appraising causality in human brain mapping studies. a** | Six criteria for appraising causality in human brain mapping studies, adapted from Bradford Hill criteria. Intracranial electrical stimulation (iES) used to illustrate each criterion. Counterfactual can be estimated by modelling a hypothetical situation in which brain activity was not modulated or was modulated differently. Specificity can be demonstrated when slightly different interventions induce measurably different outcomes. For instance, stimulating different parts of the motor cortex leads to target-specific effects on motor function. Experimental manipulation can be used to selectively modulate activity in a specific brain region compared with a control intervention (in this case a sham). Dose–response relationship evident when higher-intensity stimulation induces a higher-intensity outcome, in this case a more intense muscle contraction. Coherence can be demonstrated if different approaches converge on similar results. For instance, if stimulating a location induces a finger movement, then a lesion at the same location should induce finger weakness. Reversibility can be illustrated by reversal of behaviour when modification of brain function is discontinued. **b** | Our proposed causality continuum in human brain mapping. Example of how different brain mapping techniques may be appraised using criteria adapted from Bradford Hill criteria, ordered from least causal to most causal along a causality continuum. Targeted brain lesions and brain stimulation satisfy more causality criteria than other brain mapping techniques (described in BOX 1 and BOX 2).

Brain circuit
Any interconnected network of units, potentially including monosynaptic structural connections and/or polysynaptic functional connections.

Lesion network mapping
(LNM). A technique that identifies shared connectivity patterns among lesions associated with a specific symptom or behaviour.

symptom to specific neuroanatomy (FIG. 3a). These individual differences may be explained by damage to a common network with different components that have been affected across individuals[67,73–84] — in other words, lesions in different locations caused the same symptom because they intersected the same brain circuit[58].

To test for localization to brain circuits, lesion and stimulation data can be combined with circuit maps derived from brain imaging or electrophysiology data (BOX 3). This approach has been used widely in studies that employ lesion network mapping (LNM) and related techniques, which use correlative neuroimaging to explain heterogeneous causal findings[58]. Rather than simply comparing lesion locations, LNM first identifies the circuit connected to each lesion and then compares the circuits (FIG. 3b), as two lesions located in different brain regions may cause a similar symptom if they damage the same brain circuit[58]. In this manner, correlative neuroimaging can reveal a mechanistic connection between seemingly discordant causal results.

Because LNM maps the whole-brain connectivity of each lesion, every patient has data for every voxel in the brain, providing LNM with more statistical power than VLSM. LNM also quantifies the degree to which a lesion or stimulation site overlaps with different circuits, thus uncovering dose–response relationships[85]. Many recent studies have illustrated the potential of using normative brain connectivity to detect a causal link between lesions or stimulation sites and different clinical outcomes, even when VLSM shows no significant overlap. For instance, brain lesions are more likely to cause amnesia if they are connected to the subiculum[85]. DBS sites are more likely to relieve motor symptoms of Parkinson disease if they are connected to the supplementary motor area[86]. TMS sites are more likely to relieve depression if they are functionally anti-correlated to the subgenual cingulate[72].
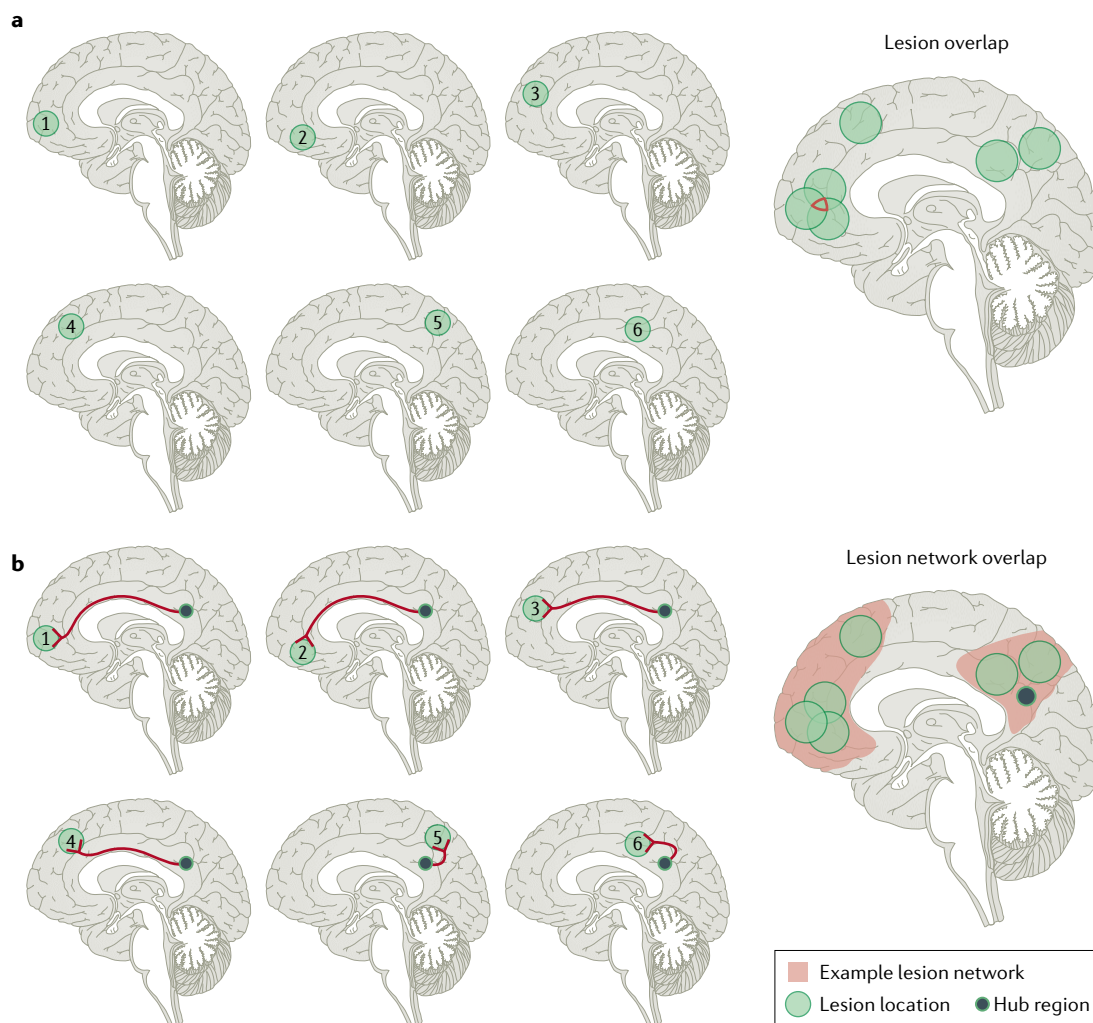
Fig. 3 | **Heterogeneous lesions causing the same symptom can complicate causal inference. a** | If lesions that cause the same symptom (but not other lesions) intersect a common brain region (area of overlap for lesions 1–3; outlined in red), one can infer a causal role of this region in symptom generation. If other lesions that cause the same symptom fail to intersect this brain region (lesions 4–6), this causal inference becomes weaker. **b** | If the locations of lesions that cause the same symptom (but not any other lesions) are connected to a common hub region (lesions 1–6), then connectivity with this hub region defines a brain network (orange) that encompasses all lesion locations causing the symptom. One can then infer a causal role of this brain network in symptom generation. Note that the hub region is not necessarily causally linked to the symptom but, rather, defines a network that is causally linked to the symptom.

Box 3 | **Circuit mapping approaches that may be combined with causal techniques**

Several imaging techniques can provide reliable information about the organization of brain circuits. These approaches can be combined with causal information from lesions or brain stimulation sites to map effects to brain circuits. Each brain mapping technique has different strengths and limitations.

Structural connectivity
Diffusion MRI can measure the locations of white matter tracts, enabling us to study the effects of stimulating different parts of the same tract[86,90,97,110]. This provides high spatial resolution, but is unable to map polysynaptic relationships or crossing white matter fibres.

Functional connectivity
This approach estimates connectivity by measuring correlations between spontaneous fluctuations of distant brain regions, usually measured with functional MRI (fMRI). Unlike diffusion MRI, this technique can estimate the effects of a stimulation site on whole-brain polysynaptic networks[75,76]. However, it is limited by the assumption that temporal correlation implies a connection between different brain regions.

Normative connectivity
Rather than measuring structural or functional connectivity in individuals, it is possible to use a large connectome database to estimate the whole-brain connectivity of a particular brain region. This technique provides greater reliability at the expense of mapping inter-individual variability in brain architecture[58].

Surface electrophysiology
Electroencephalography or magnetoencephalography can be used to measure the immediate electrical effects of stimulating distant brain regions, providing high temporal precision[14]. These techniques enable the study of the rapid propagation of the stimulation to the rest of the brain, but are limited by poor spatial resolution.

Intracranial electrophysiology
Owing to greater anatomical precision than surface electroencephalography, intracranial electroencephalography enables precise mapping of specific behaviours. However, these tools are invasive and are usually only practical in specialized clinical centres where patients are evaluated for surgical treatment of refractory epilepsy[66].

Other approaches
In theory, it may be possible to define a priori circuits using any neuroimaging approach, including task fMRI, effective connectivity, regional cerebral blood flow or volumetric imaging. These approaches have not yet been widely used to map the circuits connected to specific lesions or stimulation sites, so methodological validation is still needed.

Intracranial electrical stimulation
(iES). Invasive technique for modulating brain activity using depth electrodes or subdural electrodes, commonly used for mapping brain function prior to brain surgery.

Similar approaches have now been used to map the causal neuroanatomy of movement disorders[86–91], mood disorders[71,90,92–95], anxiety-related disorders[96–98], psychotic disorders[99–101], disorders of consciousness[102–104] and various other neuropsychiatric phenomena[105–109].

Of note, most LNM studies define the circuit using a normative connectome database, which acts as a 'wiring diagram' that depicts the expected connectivity profile of any lesion location. This wiring diagram is based on resting-state functional connectivity or diffusion tractography data from a large cohort of participants without brain lesions or stimulation. Ongoing work is underway to compare wiring diagrams generated by different methods, including structural versus functional connectivity[110,111] and normative versus individualized connectivity[75,76], each of which may independently add value[75,76,111].

There are also some disadvantages to studying lesions and stimulation sites at the circuit level. These approaches can increase the risk of erroneous causal reasoning by adding an extra layer of complexity. If a brain lesion causes a particular symptom, it is reasonable to infer that the lesion location is causally involved in the symptom. If different lesion locations causing the same symptom are connected to one common hub, it is tempting to infer a causal relationship between the symptom and the hub. However, the hub simply defines the connectivity network that encompasses the lesion locations that cause a symptom. In other words, the causal inference is for the network, not the hub that defines the network. The hub may represent an important treatment target, but it is one step removed from the causal inference applied to the lesion locations themselves. One would need to independently study the effect of lesions or stimulation to the hub region to infer a causal link between the hub and the symptoms.

Real-time causal mapping. Another approach to causal inference addresses the temporality, experimental manipulation and reversibility criteria by using real-time stimulation experiments in which different brain regions are focally stimulated to induce transient effects.

The most precise tool for real-time mapping of human brain function is intracranial electrical stimulation (iES) in conscious individuals[66], although recent studies suggest that thermal stimulation may achieve similar effects[112,113]. These studies are usually conducted in patients who are implanted with multifocal electrode arrays for seizure monitoring before surgery for epilepsy. Although these electrodes are implanted, investigators can use them to deliver focal stimulation and measure changes in a participant's subjective experiential state or experimental task performance[66,114]. This method also enables simultaneous intracranial electroencephalography recording during and after stimulation, yielding measures of effective connectivity. Although iES does not provide whole-brain coverage in the same manner as LNM and other circuit-level analyses, one can still estimate the counterfactual by comparing clinical outcomes from thousands of possible stimulation sites across multiple patients[67].

Although iES dates back to Penfield's work in the mid-twentieth century, a recent renaissance in iES research has been accelerated by the development of neuroimaging and circuit-based techniques[66]. High-resolution imaging methods can localize the stimulated brain areas with millimetre precision to calculate the extent of the brain targeted by a given electrical dose[115].

Some recent studies provide examples of how iES-induced effects can be used to validate anatomical boundaries defined by other techniques such as functional neuroimaging or even animal models. For example, visual perception of faces is modified by stimulation of a patch of the posterior fusiform face area defined by task-based fMRI[80]. A will to persevere can be induced by stimulation of a hub in the anterior cingulate cortex defined by resting-state fMRI[79]. An altered sense of self can be induced by stimulation of part of the posteromedial cortex implicated in self-referential thought from task-based fMRI[116]. Furthermore, the probability of observing any behavioural response to stimulation aligns with brain network boundaries defined by resting-state fMRI[67]. Finally, a dissociative state can be induced by stimulating the posterior cingulate cortex, a target identified based on an optogenetic mouse model

of dissociation[117]. These studies illustrate how iES can add value to results from correlative neuroimaging experiments or studies in animal models, potentially facilitating their translation into clinical treatments.

Clinical causal brain mapping

Optimizing brain stimulation targets. Causal research is uniquely valuable in medicine because of its potential to identify targets for clinical intervention. If a particular factor is causal, it affects the probability of an outcome with all else being equal. Therefore, modifying that factor should modify the probability of that outcome. In clinical neuroscience, this principle can be tested using existing data in patients who received therapeutic TMS and DBS. If a particular brain region is causally implicated in a symptom, then stimulating that region should modify the probability or severity of that symptom. For instance, TMS can be used to localize the language centres in individual patients by studying the effects of targeted inhibition on speech function[118,119]. This approach can be used by neurosurgeons to avoid damaging language function during surgery[120].

Furthermore, circuit mapping studies have used connectivity to gain insight into the most effective TMS and DBS targets. In major depression, the most effective TMS sites are functionally anti-correlated to the subgenual cingulate cortex[71,72,94], whereas the most effective DBS sites overlap with tracts that connect the subgenual cingulate to the orbitofrontal cortex and the limbic system[121,122]. In Parkinson disease, the most effective DBS sites for motor symptoms are structurally connected to the supplementary motor area and functionally anti-correlated to the primary motor cortex[86]. In obsessive-compulsive disorder, the most effective

DBS sites overlap with a tract that connects the subthalamic nucleus to the prefrontal cortex[96,97] and are functionally connected to the anterior cingulate, insula and precuneus[123]. In each case, stimulation sites might be refined to better target the therapeutic connections or to potentially avoid connections associated with DBS side effects[90]. Whether this improves clinical outcomes remains to be validated prospectively.

Beyond optimizing treatment targets for any given disorder, causal circuit mapping approaches have been used to disentangle which symptom clusters respond to different stimulation sites. In two cohorts of patients with major depression, TMS sites connected to a particular circuit were more likely to improve 'dysphoric' symptoms such as sadness and suicidality, whereas TMS sites connected to a different circuit were more likely to improve 'anxiosomatic' symptoms such as insomnia and sexual dysfunction[98]. Disentangling symptom clusters in this fashion lays the foundation for personalized target selection[124] and highlights the need for deeper phenotyping, which may reveal more distinct treatment targets for distinct behaviours or symptoms.

Identifying brain stimulation targets. This framework may be extended to identify new TMS and DBS targets based on lesions that modify different symptoms. If damage to a particular circuit can cause a symptom, it is reasonable to hypothesize that stimulation of the same circuit may relieve that symptom. In fact, a similar approach was used to identify some of the neuromodulation targets currently in clinical use. Early studies on incidental stroke lesions led to identification of the dorsolateral prefrontal cortex as an effective TMS target for depression[24,125,126], whereas early findings on therapeutic lesioning led to identification of the ventral intermediate nucleus, subthalamic nucleus and globus pallidus interna as effective DBS targets for Parkinson disease[19,20,22,127].

Broadly, these successes suggest that brain lesions can inform treatment targeting with brain stimulation. The causal circuit mapping approach also illustrates coherence across multiple methods, a rarely met criterion on the causality continuum. For example, although LNM can help us detect a causal link between lesions and outcomes based on overlap with a network, the causal agent is still the lesion location, not the network hub (FIG. 3). To determine whether the network hubs are causally implicated in the behaviour, targeted stimulation could then be used, thereby meeting the coherence criterion. This postulate is illustrated by two recent examples of coherence between LNM and iES results. First, lesions connected to the dorsal cingulate can disrupt volitional will[105], whereas iES to the dorsal cingulate can induce a will to persevere[79] (FIG. 4a). Second, lesions connected to the right posterior fusiform gyrus can disrupt facial recognition[107], whereas iES of the right posterior, but not the left or anterior, fusiform gyrus can distort face perception[80,128,129] (FIG. 4b). These findings illustrate how stimulation and lesions can provide complementary information about the function of a network hub.

A similar postulate suggests that different stimulation modalities can also provide complementary causal information. If a circuit is causally implicated in a symptom,
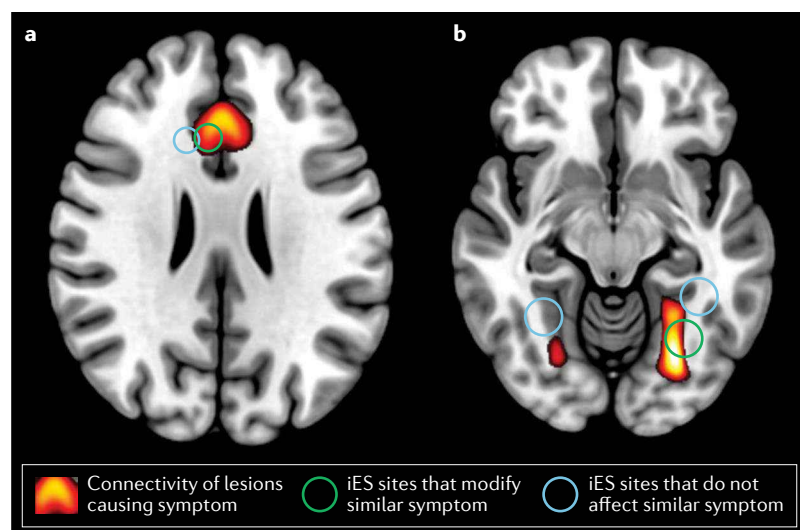
Fig. 4 | **Coherence between LNM and iES studies. a** | Lesion network mapping (LNM) revealed that lesions causing decreased will to perform actions are connected to the anterior cingulate (warm colours)[105]. Intracranial electrical stimulation (iES) sites that overlap with this network (green circle) can induce a will to persevere, unlike other nearby iES sites (blue circle)[79]. **b** | LNM revealed that lesions causing disordered face perception are connected to the right posterior fusiform gyrus (warm colours)[107]. iES to the right posterior fusiform gyrus (green circle), but not the right anterior fusiform gyrus or left fusiform gyrus (blue circles), causes distorted face perception[80,128,129]. Data for panel **a** from REFS[79,105]. Data for panel **b** from REFS[80,107,128,129].
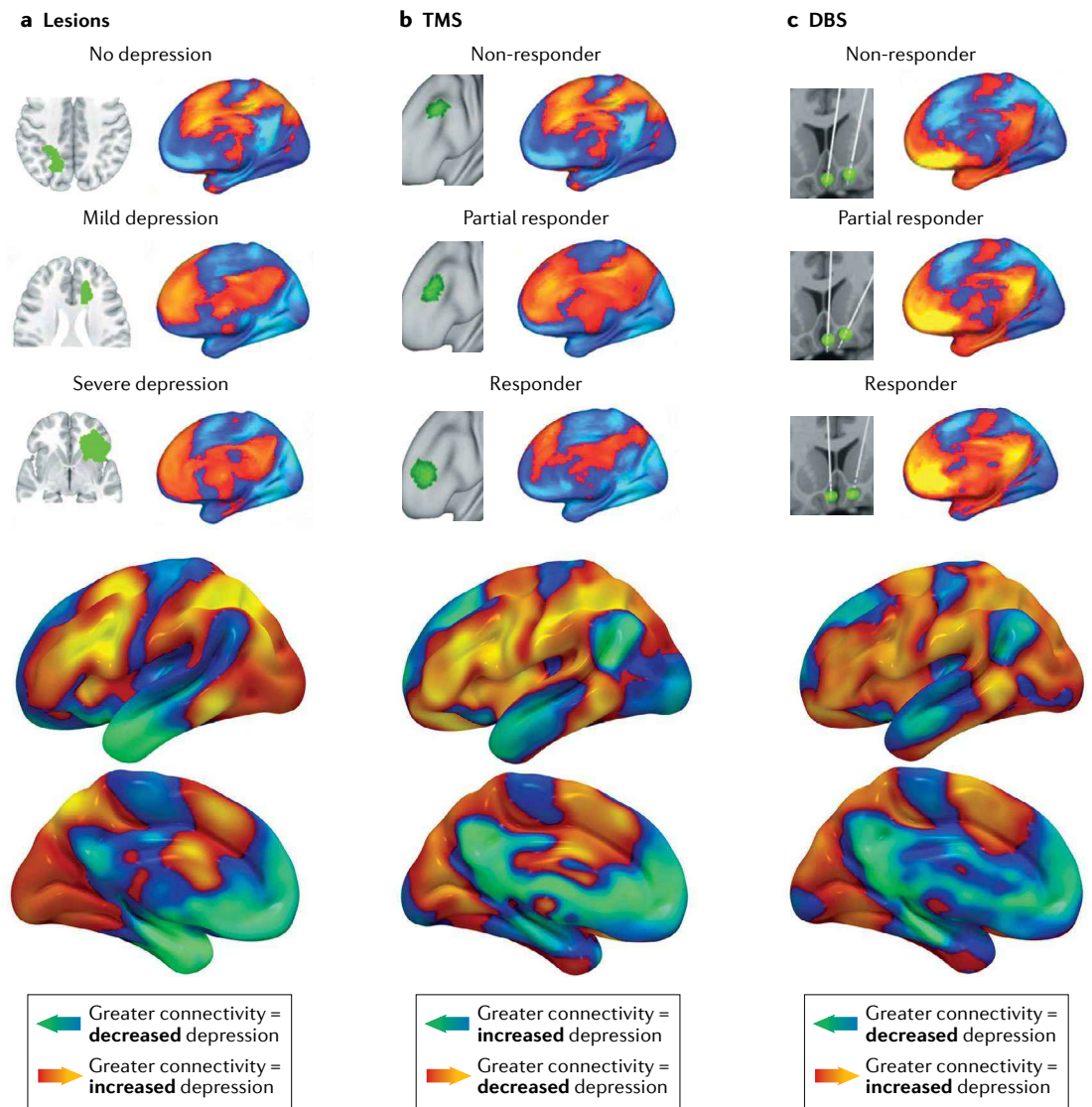
Fig. 5 | **Coherence between TMS, DBS and brain lesions.** Common brain circuit connected to brain lesions that cause depression, transcranial magnetic stimulation (TMS) sites that relieve depression and deep brain stimulation (DBS) sites that modify depression. **a** | For 461 incidental brain lesions (top left), whole-brain connectivity was estimated using a normative connectome database (top right). This connectivity map was compared with depression severity, yielding a map of the whole-brain circuit connected to lesions that increase depression severity (bottom). **b** | Same procedure conducted for 151 incidentally variable TMS sites, yielding a map of the whole-brain circuit connected to TMS sites that improve depression. Colour scale inverted for this map because TMS sites that improve depression were expected to be anti-correlated to lesion locations associated with lower depression severity[109]. **c** | Same procedure also conducted for 101 DBS sites, yielding a map of the whole-brain circuit connected to DBS sites that modify depression. These three maps are significantly more similar than expected by chance. Adapted from REF.[95], Springer Nature Ltd.

then the symptom should be affected by different types of stimulation to that circuit. A review of published TMS and DBS targets explored this topic across 14 different disorders, finding that published TMS targets for various disorders were connected to the same brain regions as published DBS targets for the same disorders[109]. The authors thus hypothesized that better DBS targets may be identified based on the connectivity of effective TMS targets, and vice versa. If TMS and DBS sites converge on the same circuit, the strength of the causal inference linking this circuit to the symptom increases.

A recent study systematically tested this hypothesis by examining the severity of depression in 461 patients with brain lesions (five datasets), 151 patients receiving therapeutic TMS (four datasets) and 101 patients receiving therapeutic DBS (five datasets)[95]. Across all 14 datasets, this analysis revealed a common circuit that was connected to the lesions that caused depression, the TMS sites that improved depression and the DBS sites that worsened or improved depression (FIG. 5). Connectivity to this circuit also predicted the clinical efficacy of TMS and DBS sites in a 'leave one dataset out' cross-validation. Similar concordance was also observed for motor symptoms of Parkinson disease — lesions that caused parkinsonism were connected to the same circuit as DBS sites that relieve parkinsonism, and both were

connected to the primary motor cortex, the most common TMS target for parkinsonism[95]. Thus, the connectivity of brain lesions that cause a symptom may reveal effective stimulation targets for the same symptom[95]. This provides a clear model for how convergent results can strengthen causal inference — if a circuit is causally linked to a symptom, then different manipulations of the circuit should modify the probability or severity of that symptom. Subsequent studies have expanded this principle to other symptoms and disorders, finding that lesion locations can predict which DBS targets are most likely to modify tics in Tourette disorder[130] and cognitive symptoms in Parkinson disease[131].

Road map for future research

Integrating correlative and causal maps. Beyond LNM, which directly combines correlative and causal techniques into a single analysis, there are several ways to test for coherence (or lack thereof) between correlative and causal techniques. This principle was nicely illustrated in two recent studies, in which task-based intracranial electroencephalography (causality score = 2) was used to identify a medial–lateral dissociation between simple tone perception and complex speech perception. These correlative results were then used to guide a causal manipulation (iES, causality score = 5.5), revealing that medial sites induced hallucinations of simple sounds whereas lateral sites interrupted complex word perception[132,133]. Thus, by combining advanced analysis of correlative data with causal stimulation, these studies discovered a novel medial–lateral gradient in the auditory cortex.

Causal techniques can also be used to compare different correlative brain mapping results. For instance, a recent study used task-based fMRI (causality score = 2) to compare different approaches for parcellating the brain based on task-free neuroimaging (causality score = 0)[134]. Another recent study used incidental lesions (causality score = 5.5) to compare different approaches for defining 'cortical hubs' based on resting-state fMRI (causality score = 0)[135]. In both cases, the more causal technique showed that some of the correlative approaches were more functionally relevant than the others. Importantly, the strength of causal inference in such studies is defined by the more causal technique, not the less causal technique. In the future, techniques further along the causality continuum should be used to test results from less causal methods whenever feasible.

By contrast, correlative imaging can also be used as a biomarker when comparing the effects of different causal techniques, especially techniques that do not induce a measurable clinical effect. For instance, functional neuroimaging can be used to estimate the effect of TMS to different targets in the prefrontal cortex, even when these targets do not induce clear behavioural changes[136–138]. In addition, by using correlative imaging to measure the physiological effects of an intervention before the clinical effects become apparent, it may be possible to derive useful therapeutic biomarkers[139].

Finally, multifocal stimulation protocols may be designed to target different parts of a circuit defined by correlative techniques. The effect of stimulating different nodes of the same circuit could differ, and stimulating one site may modify excitability of a secondary site[132]. Various studies have attempted multifocal TMS and some have suggested that it may be safe to combine TMS with DBS applied to the same circuit, but more work is needed to identify optimal multifocal stimulation approaches[140–143].

Best practices for causal brain mapping. Different human brain mapping techniques can yield insights at different levels of causal inference. It is important to recognize which level of causal inference applies to a given study and to form appropriate conclusions as a result, especially when the goal is uncovering a therapeutic target. Here, we provide some guidelines that may be helpful in this regard.

First, purely correlative studies should avoid making claims of causality and be cautious in suggesting therapeutic relevance. Although correlative studies can plausibly identify biomarkers for diagnosis and monitoring[139], there is a fundamental gap between correlative results and identifying treatment targets — whereas some correlations may reflect a causal relationship, others could be spurious or reverse-causal.

Second, the strength of causal inference should be explicitly assessed in studies that make any causal conclusions, including limitations. Before proposing a treatment target, brain mapping results should be explicitly compared with results from studies using approaches further along the causality continuum. For instance, atrophy-derived brain maps are more likely to be causal if they agree with lesion-derived maps of the same behaviour[144]. If there is concordance across different causal mapping approaches, and consistency as one moves higher up on the causality continuum, this increases the probability that a correlative finding provides a useful mechanistic insight towards a treatment target.

Third, lesion and stimulation studies should explicitly consider the degree of concordance between modalities. This may become increasingly common as more lesion and stimulation datasets become available. We provided several examples of concordance between lesion and brain stimulation data[95,130,131] (FIGS 4,5), which increases confidence in a causal relationship.

Fourth, to identify treatment targets, it is useful to focus not only on the strength of causality but also on its clinical relevance. For instance, if stimulating a specific region or circuit modifies the probability of specific symptoms, this is clinically meaningful and may represent a therapeutic target. By contrast, if stimulating a specific region or circuit induces an electrophysiological or haemodynamic response in a distant region, this information may be less clinically relevant. This knowledge may be useful for mechanistic and methodological purposes, but it does not imply that the distant region should be used as a treatment target.

Fifth, there is a critical difference between the strength of causality and the overall quality of a study. Our proposed causality continuum applies to studies that are well designed and well executed with rigorous methods and appropriate controls. If a study uses unjustified methods or lacks an appropriate control group,

then the strength of causal inference cannot be reliably assessed, regardless of which methods were used. Thus, the causality continuum is only one of many dimensions that should be assessed when appraising a human brain mapping study.

Last, and perhaps most important, prospective randomized clinical trials are still needed to translate causal circuit mapping results into practical treatments. Most existing clinical trials compare the effect of treatment with a sham treatment or placebo. This approach is valuable to determine the effect of the treatment modality, but does not clarify the effect of the targeted circuit. To address this, prospective trials could isolate the effect of the targeted circuit by comparing the clinical effects of stimulating different brain circuits. If different targets can modulate different symptoms, these trials would be a powerful tool for testing circuit-based hypotheses and translating these findings into treatments.

Limitations and open questions. The mechanisms by which lesions or stimulations change an outcome remain unclear. It is relatively straightforward to understand why brain damage from a structural lesion will cause a loss of function, but it is unclear how some lesions can lead to positive outcomes[89,145]. Lesions are generally presumed to be inhibitory, but homeostatic compensation and diaschisis may lead to excitatory effects elsewhere the brain[146,147]. The effects of focal brain damage may extend beyond the lesion itself and these effects may differ for grey matter versus white matter lesions[147,148]. A better understanding of the neural mechanisms of diaschisis may shed light on these open questions.

The mechanisms of transcranial and intracranial stimulation can add complexity to the interpretation of causal brain mapping studies, as the roles of different stimulation parameters remain unclear. What are the effects of different stimulation frequencies and intensities[149,150]? How are different cortical layers affected[151]? How are outcomes related to cognitive states or tasks[152]? These and other factors may influence which neurons are affected by a stimulation[153,154]. Subtle variations in the stimulated field can lead to measurable differences in behavioural outcomes[155], which may explain why some of Penfield's experiments have failed to replicate[156].

Results have been particularly heterogeneous in the TMS literature. This is partly related to dosing — over the past two decades, TMS studies have used progressively higher doses and observed progressively larger effects[157]. However, individual differences and mechanisms remain poorly understood. For example, high-frequency TMS is conventionally thought to increase cortical excitability, but up to 30% of individuals may show no change in, or even a decrease in, excitability[158]. Similarly, TMS has been hypothesized to function by modulating NMDA receptor-mediated plasticity, which can be modified by various drugs[159], but these drugs can have minimal effect on TMS clinical outcomes[160]. As such, even a rigorously designed TMS study may produce limited effect sizes, potentially underestimating the true causal relationship between the targeted brain region and behavioural effect.

Finally, when attempting to map causal information to brain circuits, it is important to note that brain circuits can be defined in different ways[5,161–164]. Human brain circuits can be defined using thousands of individual voxels[63,100,165], hundreds of distinct parcels[5,161,163,164] or a handful of functionally independent networks[6,162], and can be defined at the group or individual level[166–169]. This heterogeneity leads to high researcher degrees of freedom and may contribute to poor reproducibility of neuroimaging results[170]. This fragmentation was illustrated by a recent study in which 70 different research teams were asked to test the same hypotheses in the same neuroimaging dataset, and no two teams chose the same analytical procedure[171]. Although many causal mapping studies have been robust to subtle methodological variations[93,98], others have yielded different results when using different approaches[172]. Methodological standardization may improve reproducibility (one of Bradford Hill's criteria), and thus help move the entire field closer to causal inference.

Conclusions

Owing to decades of research in clinical and systems neuroscience, the field has amassed a wealth of knowledge about the organization of different brain regions. The recent re-emergence of lesion and stimulation-based studies has also led to a rapid growth in our ability to infer the causal role of these different regions, potentially enabling more informed clinical trials of targeted neuromodulation. To achieve this, causality should be interpreted as 'stronger' or 'weaker' along a continuous spectrum, and purely correlative findings should be explicitly acknowledged as such. By using tools and approaches that are further along the causality continuum, we can come closer to translating brain maps into clinical treatment targets for neuropsychiatric disease. Following the example set by the field of econometrics[15], this may help human brain mapping achieve its own credibility revolution.

1. Fox, M. D. & Raichle, M. E. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.* **8**, 700–711 (2007).
2. Raichle, M. E. A brief history of human brain mapping. *Trends Neurosci.* **32**, 118–126 (2009).
3. Genon, S., Reid, A., Langner, R., Amunts, K. & Eickhoff, S. B. How to characterize the function of a brain region. *Trends Cogn. Sci.* **22**, 350–364 (2018).
4. Gusnard, D. A. & Raichle, M. E. Searching for a baseline: functional imaging and the resting human brain. *Nat. Rev. Neurosci.* **2**, 685–694 (2001).
5. Glasser, M. F. et al. A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
6. Power, J. D. et al. Functional network organization of the human brain. *Neuron* **72**, 665–678 (2011).
7. Bullmore, E. & Sporns, O. The economy of brain network organization. *Nat. Rev. Neurosci.* **13**, 336–349 (2012).
8. Logothetis, N. K. What we can do and what we cannot do with fMRI. *Nature* **453**, 869–878 (2008).
   This work presents an overview of common uses and pitfalls of correlative neuroimaging.
9. Jonas, E. & Kording, K. P. Could a neuroscientist understand a microprocessor? *PLoS Comput. Biol.* **13**, e1005268 (2017).
   This work illustrates that conventional neuroscience tools are unable to explain the function of a microprocessor (a device that we understand completely), raising questions about whether they can explain anything meaningful about the brain.
10. Reid, A. T. et al. Advancing functional connectivity research from association to causation. *Nat. Neurosci.* **22**, 1751–1760 (2019).
11. Laumann, T. O. & Snyder, A. Z. Brain activity is not only for thinking. *Curr. Opin. Behav. Sci.* **40**, 130–136 (2021).

12. Etkin, A. A reckoning and research agenda for neuroimaging in psychiatry. *Am. J. Psychiatry* **176**, 507–511 (2019).

13. Weichwald, S. & Peters, J. Causality in cognitive neuroscience: concepts, challenges, and distributional robustness. *J. Cogn. Neurosci.* **33**, 226–247 (2021).

14. Etkin, A. Addressing the causality gap in human psychiatric neuroscience. *JAMA Psychiatry* **75**, 3–4 (2018).

15. Angrist, J. D. & Pischke, J.-S. The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. *J. Econ. Perspect.* **24**, 3–30 (2010).

16. Leamer, E. E. Let's take the con out of econometrics. *Am. Econ. Rev.* **73**, 31–43 (1983).

17. Etkin, A. Mapping causal circuitry in human depression. *Biol. Psychiatry* **86**, 732–733 (2019).

18. Bergmann, T. O. & Hartwigsen, G. Inferring causality from noninvasive brain stimulation in cognitive neuroscience. *J. Cogn. Neurosci.* **33**, 195–225 (2021).

19. Pycroft, L., Stein, J. & Aziz, T. Deep brain stimulation: an overview of history, methods, and future developments. *Brain Neurosci. Adv.* **2**, 2398212818816017 (2018).

20. Lozano, A. M. et al. Effect of GPi pallidotomy on motor function in Parkinson's disease. *Lancet* **346**, 1383–1387 (1995).

21. Groiss, S. J., Wojtecki, L., Sudmeyer, M. & Schnitzler, A. Deep brain stimulation in Parkinson's disease. *Ther. Adv. Neurol. Disord.* **2**, 20–28 (2009).

22. Benabid, A. L. et al. Long-term suppression of tremor by chronic stimulation of the ventral intermediate thalamic nucleus. *Lancet* **337**, 403–406 (1991).
   This work is an early example of successful therapeutic use of DBS, using a target for tremor based partly on prior studies using brain lesions.

23. George, M. S. et al. A controlled trial of daily left prefrontal cortex TMS for treating depression. *Biol. Psychiatry* **48**, 962–970 (2000).

24. Pascual-Leone, A., Rubio, B., Pallardó, F. & Catalá, M. D. Rapid-rate transcranial magnetic stimulation of left dorsolateral prefrontal cortex in drug-resistant depression. *Lancet* **348**, 233–237 (1996).
   This example of successful therapeutic use of TMS uses a target for depression based, in part, on depression-causing lesions — later confirmed in larger trials, including.

25. Harlow, J. M. Passage of an iron rod through the head. *Boston Med. Surg. J.* **39**, 389–393 (1848).

26. Scoville, W. B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* **20**, 11 (1957).

27. Penfield, W. & Jasper, H. *Epilepsy and the Functional Anatomy of the Human Brain* (Little, Brown and Co., 1954).

28. Penfield, W. & Perot, P. The brain's record of auditory and visual experience: a final summary and discussion. *Brain* **86**, 595–696 (1963).

29. Penfield, W. & Boldrey, E. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain* **60**, 389–443 (1937).

30. Rorden, C. & Karnath, H.-O. Using human brain lesions to infer function: a relic from a past era in the fMRI age? *Nat. Rev. Neurosci.* **5**, 812–819 (2004).

31. Michel, J.-B. et al. Quantitative analysis of culture using millions of digitized books. *Science* **331**, 176–182 (2011).

32. Gaillard, R. et al. Direct intracranial, FMRI, and lesion evidence for the causal role of left inferotemporal cortex in reading. *Neuron* **50**, 191–204 (2006).

33. Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M. & Raichle, M. E. Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature* **331**, 585–589 (1988).
   This classic study uses correlative neuroimaging to illustrate features of language processing in healthy individuals.

34. Dolan, R. J. & Fletcher, P. Dissociating prefrontal and hippocampal function in episodic memory encoding. *Nature* **388**, 582–585 (1997).

35. Zalesky, A., Fornito, A. & Bullmore, E. On the use of correlation as a measure of network connectivity. *Neuroimage* **60**, 2096–2106 (2012).

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

37. Pearl, J. *Causality* (Cambridge Univ. Press, 2009).

38. Angrist, J. D. & Pischke, J.-S. *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton Univ. Press, 2008).

39. Holland, P. W. Statistics and causal inference. *J. Am. Stat. Assoc.* **81**, 945–960 (1986).

40. Pearl, J. & Mackenzie, D. *The Book of Why: The New Science of Cause and Effect* (Basic Books, 2018).
   This book details the basic principles of causal reasoning across disciplines.

41. Gotthelf, A. *Aristotle's Conception of Final Causality* Thesis (Columbia Univ., 1975).

42. Black, D. L. Mental existence in Thomas Aquinas and Avicenna. *Mediaev. Stud.* **61**, 45–79 (1999).

43. Taylor, R. in *The Routledge Companion to Islamic Philosophy* (eds Taylor, R. C. & Lopez-Farjeat, L. X.) 225–235 (Routledge, 2016).

44. Smith, N. K. *The Philosophy of David Hume: A Critical Study of Its Origins and Central Doctrines* (Macmillan, 1941).

45. Koch, R. An address on bacteriological research. *BMJ* **2**, 380 (1890).

46. Bradford Hill, A. The environment and disease: association or causation? *Proc. R. Soc. Med.* **58**, 295–300 (1965).
   This work introduces the Bradford Hill criteria, which have been used to understand causal relationships across multiple disciplines.

47. Fedak, K. M., Bernal, A., Capshaw, Z. A. & Gross, S. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. *Emerg. Themes Epidemiol.* **12**, 14 (2015).

48. Araújo, L., Dalgalarrondo, P. & Banzato, C. On the notion of causality in medicine: addressing Austin Bradford Hill and John L. Mackie. *Arch. Clin. Psychiatry* **41**, 56–61 (2014).

49. Sajadi, M. M., Mansouri, D. & Sajadi, M. R. Ibn Sina and the clinical trial. *Ann. Intern. Med.* **150**, 640–643 (2009).

50. Bradford Hill, A. Memories of the British streptomycin trial in tuberculosis: the first randomized clinical trial. *Controlled Clin. Trials* **11**, 77–79 (1990).

51. Gillies, D. *Causality, Probability and Medicine* (Taylor & Francis, 2018).

52. Kaplan, R. M. & Irvin, V. L. Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLoS ONE* **10**, e0132382 (2015).

53. Baluku, J. B. et al. Prevalence of malaria and TB coinfection at a national tuberculosis treatment centre in Uganda. *J. Tropical Med.* **2019**, 3741294 (2019).

54. Koch, R. The etiology of tuberculosis. *Mitth. aus dem Kaiserlichen Gesundheitsamte* **2**, 1–88 (1884).

55. Hernán, M. A. & Robins, J. M. Using big data to emulate a target trial when a randomized trial is not available. *Am. J. Epidemiol.* **183**, 758–764 (2016).

56. Marinescu, I. E., Lawlor, P. N. & Kording, K. P. Quasi-experimental causality in neuroscience and behavioural research. *Nat. Hum. Behav.* **2**, 891–898 (2018).
   This work presents an overview of different approaches to investigating causality when a prospective controlled experiment is not possible.

57. Cook, T. D., Shadish, W. R. & Wong, V. C. Three conditions under which experiments and observational studies produce comparable causal estimates: new findings from within-study comparisons. *J. Policy Anal. Manag.* **27**, 724–750 (2008).

58. Fox, M. D. Mapping symptoms to brain networks with the human connectome. *N. Engl. J. Med.* **379**, 2237–2245 (2018).
   This work is an introduction to LNM, which can be used to relate different lesion locations causing a symptom to common brain circuits.

59. Williamson, J. Establishing causal claims in medicine. *Int. Stud. Philos. Sci.* **32**, 33–61 (2019).

60. Russo, F. & Williamson, J. Interpreting causality in the health sciences. *Int. Stud. Philos. Sci.* **21**, 157–170 (2007).

61. Evans, A. S. Causation and disease: the Henle–Koch postulates revisited. *Yale J. Biol. Med.* **49**, 175–195 (1976).

62. Cassidy, J. M., Mark, J. I. & Cramer, S. C. Functional connectivity drives stroke recovery: shifting the paradigm from correlation to causation. *Brain* https://doi.org/10.1093/brain/awab469 (2021).

63. Bates, E. et al. Voxel-based lesion–symptom mapping. *Nat. Neurosci.* **6**, 448–450 (2003).

64. Wu, O. et al. Role of acute lesion topography in initial ischemic stroke severity and long-term functional outcomes. *Stroke* **46**, 2438–2444 (2015).

65. Corbetta, M. et al. Common behavioral clusters and subcortical anatomy in stroke. *Neuron* **85**, 927–941 (2015).

66. Parvizi, J. & Kastner, S. Promises and limitations of human intracranial electroencephalography. *Nat. Neurosci.* **21**, 474–483 (2018).

67. Fox, K. C. R. et al. Intrinsic network architecture predicts the effects elicited by intracranial electrical stimulation of the human brain. *Nat. Hum. Behav.* **4**, 1039–1052 (2020).
   This work illustrates how iES can be used to investigate the causal relevance of networks defined using task-free neuroimaging.

68. Koenigs, M. et al. Focal brain damage protects against post-traumatic stress disorder in combat veterans. *Nat. Neurosci.* **11**, 232–237 (2008).

69. Knutson, K. M. et al. Injured brain regions associated with anxiety in Vietnam veterans. *Neuropsychologia* **51**, 686–694 (2013).

70. Herbsman, T. et al. More lateral and anterior prefrontal coil location is associated with better repetitive transcranial magnetic stimulation antidepressant response. *Biol. Psychiatry* **66**, 509–515 (2009).

71. Fox, M. D., Buckner, R. L., White, M. P., Greicius, M. D. & Pascual-Leone, A. Efficacy of transcranial magnetic stimulation targets for depression is related to intrinsic functional connectivity with the subgenual cingulate. *Biol. Psychiatry* **72**, 595–603 (2012).

72. Weigand, A. et al. Prospective validation that subgenual connectivity predicts antidepressant efficacy of transcranial magnetic stimulation sites. *Biol. Psychiatry* **84**, 28–37 (2018).

73. Gordon, E. M. et al. Individual-specific features of brain systems identified with resting state functional correlations. *Neuroimage* **146**, 918–939 (2017).

74. Gordon, E. M., Laumann, T. O., Adeyemo, B. & Petersen, S. E. Individual variability of the system-level organization of the human brain. *Cereb. Cortex* **27**, 386–399 (2017).

75. Siddiqi, S. H., Weigand, A., Pascual-Leone, A. & Fox, M. D. Identification of personalized TMS targets based on subgenual cingulate connectivity: an independent replication. *Biol. Psychiatry* **90**, e55–e56 (2021).

76. Cash, R. F. H., Cocchi, L., Lv, J., Fitzgerald, P. B. & Zalesky, A. Functional magnetic resonance imaging-guided personalization of transcranial magnetic stimulation treatment for depression. *JAMA Psychiatry* **78**, 337–339 (2020).

77. Howell, B. et al. Quantifying the axonal pathways directly stimulated in therapeutic subcallosal cingulate deep brain stimulation. *Hum. Brain Mapp.* **40**, 889–903 (2019).

78. Siddiqi, S. H. et al. Repetitive transcranial magnetic stimulation with resting-state network targeting for treatment-resistant depression in traumatic brain injury: a randomized, controlled, double-blinded pilot study. *J. Neurotrauma* **36**, 1361–1374 (2019).

79. Parvizi, J., Rangarajan, V., Shirer, W. R., Desai, N. & Greicius, M. D. The will to persevere induced by electrical stimulation of the human cingulate gyrus. *Neuron* **80**, 1359–1367 (2013).

80. Schrouff, J. et al. Fast temporal dynamics and causal relevance of face processing in the human temporal cortex. *Nat. Commun.* **11**, 656 (2020).

81. Yih, J., Beam, D. E., Fox, K. C. R. & Parvizi, J. Intensity of affective experience is modulated by magnitude of intracranial electrical stimulation in human orbitofrontal, cingulate and insular cortices. *Soc. Cogn. Affect. Neurosci.* **14**, 339–351 (2019).

82. Downar, J. et al. Anhedonia and reward-circuit connectivity distinguish nonresponders from responders to dorsomedial prefrontal repetitive transcranial magnetic stimulation in major depression. *Biol. Psychiatry* **76**, 176–185 (2014).

83. Drysdale, A. T. et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat. Med.* **12**, 28–38 (2016).

84. Cash, R. F. H. et al. Using brain imaging to improve spatial targeting of transcranial magnetic stimulation for depression. *Biol. Psychiatry* **90**, 689–700 (2020).

85. Ferguson, M. A. et al. A human memory circuit derived from brain lesions causing amnesia. *Nat. Commun.* **10**, 3497 (2019).

86. Horn, A. et al. Connectivity predicts deep brain stimulation outcome in Parkinson disease. *Ann. Neurol.* **82**, 67–78 (2017).

87. Joutsa, J., Horn, A., Hsu, J. & Fox, M. D. Localizing parkinsonism based on focal brain lesions. *Brain* **141**, 2445–2456 (2018).

88. Corp, D. T. et al. Network localization of cervical dystonia based on causal brain lesions. *Brain* **142**, 1660–1674 (2019).

89. Joutsa, J. et al. Identifying therapeutic targets from spontaneous beneficial brain lesions. *Ann. Neurol.* **84**, 153–157 (2018).

90. Irmen, F. et al. Left prefrontal impact links subthalamic stimulation with depressive symptoms. *Ann. Neurol.* **87**, 962–975 (2020).
91. Laganiere, S., Boes, A. D. & Fox, M. D. Network localization of hemichorea-hemiballismus. *Neurology* **86**, 2187–2195 (2016).
92. Padmanabhan, J. L. et al. A human depression circuit derived from focal brain lesions. *Biol. Psychiatry* **86**, 749–758 (2019).
93. Cotovio, G. et al. Mapping mania symptoms based on focal brain damage. *J. Clin. Invest.* **130**, 5209–5222 (2020).
94. Cash, R. F. H. et al. Subgenual functional connectivity predicts antidepressant treatment response to transcranial magnetic stimulation: independent validation and evaluation of personalization. *Biol. Psychiatry* **86**, e5–e7 (2019).
95. Siddiqi, S. H. et al. Brain stimulation and brain lesions converge on common causal circuits in neuropsychiatric disease. *Nat. Hum. Behav.* **5**, 1707–1716 (2021).
    This work is a recent test of whether different causal sources of information (lesion locations, non-invasive brain stimulation sites and invasive brain stimulation sites) map to a common brain circuit.
96. Baldermann, J. C. et al. Connectivity profile predictive of effective deep brain stimulation in obsessive-compulsive disorder. *Biol. Psychiatry* **85**, 735–743 (2019).
97. Li, N. et al. A unified connectomic target for deep brain stimulation in obsessive-compulsive disorder. *Nat. Commun.* **11**, 3364 (2020).
98. Siddiqi, S. H. et al. Distinct symptom-specific treatment targets for circuit-based neuromodulation. *Am. J. Psychiatry* **177**, 435–446 (2020).
99. Kim, N. Y. et al. Lesions causing hallucinations localize to one common brain network. *Mol. Psychiatry* **26**, 1299–1309 (2021).
100. Boes, A. D. et al. Network localization of neurological symptoms from focal brain lesions. *Brain* **138**, 3061–3075 (2015).
101. Darby, R. R., Laganiere, S., Pascual-Leone, A., Prasad, S. & Fox, M. D. Finding the imposter: brain connectivity of lesions causing delusional misidentifications. *Brain* **140**, 497–507 (2017).
102. Snider, S. B. et al. Cortical lesions causing loss of consciousness are anticorrelated with the dorsal brainstem. *Hum. Brain Mapp.* **41**, 1520–1531 (2020).
103. Fischer, D. B. et al. A human brain network derived from coma-causing brainstem lesions. *Neurology* **87**, 2427–2434 (2016).
104. Joutsa, J., Shih, L. C. & Fox, M. D. Mapping Holmes tremor circuit using the human brain connectome. *Ann. Neurol.* **86**, 812–820 (2019).
105. Darby, R. R., Joutsa, J., Burke, M. J. & Fox, M. D. Lesion network localization of free will. *Proc. Natl Acad. Sci. USA* **115**, 10792–10797 (2018).
106. Darby, R. R., Horn, A., Cushman, F. & Fox, M. D. Lesion network localization of criminal behavior. *Proc. Natl Acad. Sci. USA* **115**, 601–606 (2018).
107. Cohen, A. L. et al. Looking beyond the face area: lesion network mapping of prosopagnosia. *Brain* **142**, 3975–3990 (2019).
108. Cohen, A. L. et al. Tuber locations associated with infantile spasms map to a common brain network. *Ann. Neurol.* **89**, 726–739 (2021).
109. Fox, M. D. et al. Resting-state networks link invasive and noninvasive brain stimulation across diverse psychiatric and neurological diseases. *Proc. Natl Acad. Sci. USA* **111**, E4367–E4375 (2014).
110. Horn, A. & Fox, M. D. Opportunities of connectomic neuromodulation. *Neuroimage* **221**, 117180 (2020).
111. Bowren, M. et al. Post-stroke cognitive and motor outcomes predicted from lesion location and lesion network mapping. *Brain* https://doi.org/10.1093/brain/awac010 (2022).
112. Long, M. A. et al. Functional segregation of cortical regions underlying speech timing and articulation. *Neuron* **89**, 1187–1193 (2016).
113. Ibayashi, K. et al. Focal cortical surface cooling is a novel and safe method for intraoperative functional brain mapping. *World Neurosurg.* **147**, e118–e129 (2021).
114. Borchers, S., Himmelbach, M., Logothetis, N. & Karnath, H. O. Direct electrical stimulation of human cortex — the gold standard for mapping brain functions? *Nat. Rev. Neurosci.* **13**, 63–70 (2011).
115. Winawer, J. & Parvizi, J. Linking electrical stimulation of human primary visual cortex, size of affected cortical area, neuronal responses, and subjective experience. *Neuron* **92**, 1213–1219 (2016).
116. Parvizi, J. et al. Altered sense of self during seizures in the posteromedial cortex. *Proc. Natl Acad. Sci. USA* **118**, e2100522118 (2021).
117. Vesuna, S. et al. Deep posteromedial cortical rhythm in dissociation. *Nature* **586**, 87–94 (2020).
118. Knecht, S. et al. Degree of language lateralization determines susceptibility to unilateral brain lesions. *Nat. Neurosci.* **5**, 695–699 (2002).
119. Pascual-Leone, A., Gates, J. R. & Dhuna, A. Induction of speech arrest and counting errors with rapid-rate transcranial magnetic stimulation. *Neurology* **41**, 697–702 (1991).
120. Lefaucheur, J. P. & Picht, T. The value of preoperative functional cortical mapping using navigated TMS. *Neurophysiol. Clin.* **46**, 125–133 (2016).
121. Riva-Posse, P. et al. Defining critical white matter pathways mediating successful subcallosal cingulate deep brain stimulation for treatment-resistant depression. *Biol. Psychiatry* **76**, 963–969 (2014).
122. Riva-Posse, P. et al. A connectomic approach for subcallosal cingulate deep brain stimulation surgery: prospective targeting in treatment-resistant depression. *Mol. Psychiatry* **23**, 843–849 (2018).
123. Li, N. et al. A unified functional network target for deep brain stimulation in obsessive-compulsive disorder. *Biol. Psychiatry* **90**, 701–713 (2021).
124. Nestor, S. M. & Blumberger, D. M. Mapping symptom clusters to circuits: toward personalizing TMS targets to improve treatment outcomes in depression. *Am. J. Psychiatry* **177**, 373–375 (2020).
125. Robinson, R. G., Kubos, K. L., Starr, L. B., Rao, K. & Price, T. R. Mood disorders in stroke patients. Importance of location of lesion. *Brain* **107**, 81–93 (1984).
126. Robinson, R. G. & Szetela, B. Mood change following left hemispheric brain injury. *Ann. Neurol.* **9**, 447–453 (1981).
127. Bergman, H., Wichmann, T. & DeLong, M. R. Reversal of experimental parkinsonism by lesions of the subthalamic nucleus. *Science* **249**, 1436–1438 (1990).
128. Parvizi, J. et al. Electrical stimulation of human fusiform face-selective regions distorts face perception. *J. Neurosci.* **32**, 14915–14920 (2012).
129. Rangarajan, V. et al. Electrical stimulation of the left and right human fusiform gyrus causes different effects in conscious face perception. *J. Neurosci.* **34**, 12828–12836 (2014).
130. Ganos, C. et al. A neural network for tics: insights from causal brain lesions and deep brain stimulation. *Brain* https://doi.org/10.1093/brain/awac009 (2022).
131. Reich, M. M. et al. A brain network for deep brain stimulation induced cognitive decline in Parkinson's disease. *Brain* https://doi.org/10.1093/brain/awac012 (2022).
132. Hamilton, L. S., Oganian, Y., Hall, J. & Chang, E. F. Parallel and distributed encoding of speech across human auditory cortex. *Cell* **184**, 4626–4639.e13 (2021).
133. Hamilton, L. S., Edwards, E. & Chang, E. F. A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. *Curr. Biol.* **28**, 1860–1871.e4 (2018).
134. Zhi, D., King, M. & Diedrichsen, J. Evaluating brain parcellations using the distance controlled boundary coefficient. Preprint at *bioRxiv* https://doi.org/10.1101/2021.05.11.443151 (2021).
135. Warren, D. E. et al. Network measures predict neuropsychological outcome after brain injury. *Proc. Natl Acad. Sci. USA* **111**, 14247–14252 (2014).
136. Eldaief, M. C., Halko, M. A., Buckner, R. L. & Pascual-Leone, A. Transcranial magnetic stimulation modulates the brain's intrinsic activity in a frequency-dependent manner. *Proc. Natl Acad. Sci. USA* **108**, 21229–21234 (2011).
137. Chen, A. C. et al. Causal interactions between fronto-parietal central executive and default-mode networks in humans. *Proc. Natl Acad. Sci. USA* **110**, 19944–19949 (2013).
138. Ozdemir, R. A. et al. Individualized perturbation of the human connectome reveals reproducible biomarkers of network dynamics relevant to cognition. *Proc. Natl Acad. Sci. USA* **117**, 8115–8125 (2020).
139. Gallen, C. L. & D'Esposito, M. Brain modularity: a biomarker of intervention-related plasticity. *Trends Cogn. Sci.* **23**, 293–304 (2019).
140. Chiu, D. et al. Multifocal transcranial stimulation in chronic ischemic stroke: a phase 1/2a randomized trial. *J. Stroke Cerebrovasc. Dis.* **29**, 104816 (2020).
141. Brys, M. et al. Multifocal repetitive TMS for motor and mood symptoms of Parkinson disease: a randomized trial. *Neurology* **87**, 1907–1915 (2016).
142. Magsood, H., Syeda, F., Holloway, K., Carmona, I. C. & Hadimani, R. L. Safety study of combination treatment: deep brain stimulation and transcranial magnetic stimulation. *Front. Hum. Neurosci.* **14**, 123 (2020).
143. Deng, Z., Lisanby, S. H. & Peterchev, A. V. in *Annual Int. Conf. IEEE Engineering in Medicine and Biology* 6821-6824 (IEEE, 2010).
144. Tetreault, A. M. et al. Network localization of clinical, cognitive, and neuropsychiatric symptoms in Alzheimer's disease. *Brain* **143**, 1249–1260 (2020).
145. Kapur, N. Paradoxical functional facilitation in brain-behaviour research. A critical review. *Brain* **119**, 1775–1790 (1996).
146. Carrera, E. & Tononi, G. Diaschisis: past, present, future. *Brain* **137**, 2408–2422 (2014).
147. Sperber, C. Rethinking causality and data complexity in brain lesion-behaviour inference and its implications for lesion-behaviour modelling. *Cortex* **126**, 49–62 (2020).
148. Griffis, J. C., Metcalf, N. V., Corbetta, M. & Shulman, G. L. Structural disconnections explain brain network dysfunction after stroke. *Cell Rep.* **28**, 2527–2540.e9 (2019).
149. Klomjai, W., Katz, R. & Lackmy-Vallee, A. Basic principles of transcranial magnetic stimulation (TMS) and repetitive TMS (rTMS). *Ann. Phys. Rehabil. Med.* **58**, 208–213 (2015).
150. Su, D. et al. Frequency-dependent effects of subthalamic deep brain stimulation on motor symptoms in Parkinson's disease: a meta-analysis of controlled trials. *Sci. Rep.* **8**, 14456 (2018).
151. Radman, T., Ramos, R. L., Brumberg, J. C. & Bikson, M. Role of cortical cell type and morphology in subthreshold and suprathreshold uniform electric field stimulation in vitro. *Brain Stimul.* **2**, 215–228 (2009).
152. Feredoes, E., Heinen, K., Weiskopf, N., Ruff, C. & Driver, J. Causal evidence for frontal involvement in memory target maintenance by posterior brain areas during distracter interference of visual working memory. *Proc. Natl Acad. Sci. USA* **108**, 17510–17515 (2011).
153. Duffley, G., Anderson, D. N., Vorwerk, J., Dorval, A. D. & Butson, C. R. Evaluation of methodologies for computing the deep brain stimulation volume of tissue activated. *J. Neural Eng.* **16**, 066024 (2019).
154. Thielscher, A., Antunes, A. & Saturnino, G. B. Field modeling for transcranial magnetic stimulation: a useful tool to understand the physiological effects of TMS? *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2015**, 222–225 (2015).
155. Raij, T. et al. Prefrontal cortex stimulation enhances fear extinction memory in humans. *Biol. Psychiatry* **84**, 129–137 (2018).
156. Curot, J. et al. Awake craniotomy and memory induction through electrical stimulation: why are Penfield's findings not replicated in the modern era? *Neurosurgery* **87**, E130–E137 (2020).
157. Cole, E. J. et al. Stanford accelerated intelligent neuromodulation therapy for treatment-resistant depression. *Am. J. Psychiatry* **177**, 716–726 (2020).
158. Maeda, F., Keenan, J. P., Tormos, J. M., Topka, H. & Pascual-Leone, A. Interindividual variability of the modulatory effects of repetitive transcranial magnetic stimulation on cortical excitability. *Exp. Brain Res.* **133**, 425–430 (2000).
159. Brown, J. C. et al. NMDA receptor partial agonist, d-cycloserine, enhances 10 Hz rTMS-induced motor plasticity, suggesting long-term potentiation (LTP) as underlying mechanism. *Brain Stimul.* **13**, 530–532 (2020).
160. Minzenberg, M. J. & Leuchter, A. F. The effect of psychotropic drugs on cortical excitability and plasticity measured with transcranial magnetic stimulation: implications for psychiatric treatment. *J. Affect. Disord.* **253**, 126–140 (2019).
161. Schaefer, A. et al. Local–global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
162. Yeo, B. T. et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
163. Gordon, E. M. et al. Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb. Cortex* **26**, 288–303 (2016).

164. Destrieux, C., Fischl, B., Dale, A. & Halgren, E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* **53**, 1–15 (2010).

165. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. & Wager, T. D. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* **8**, 665–670 (2011).

166. Laumann, T. O. et al. Functional system and areal organization of a highly sampled individual human brain. *Neuron* **87**, 657–670 (2015).

167. Gordon, E. M. et al. Precision functional mapping of individual human brains. *Neuron* **95**, 791–807.e7 (2017).

168. Wang, D. et al. Parcellating cortical functional networks in individuals. *Nat. Neurosci.* **18**, 1853–1860 (2015).

169. Hacker, C. D. et al. Resting state network estimation in individual subjects. *Neuroimage* **82**, 616–633 (2013).

170. Cohen, A. L. & Fox, M. D. Reply: The influence of sample size and arbitrary statistical thresholds in lesion-network mapping. *Brain* **143**, e41 (2020).

171. Botvinik-Nezer, R. et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* **582**, 84–88 (2020).

172. Salvalaggio, A. et al. Reply: Lesion network mapping predicts post-stroke behavioural deficits and improves localization. *Brain* **144**, e36 (2021).

173. Bergmann, T. O. et al. Concurrent TMS–fMRI for causal network perturbation and proof of target engagement. *Neuroimage* **237**, 118093 (2021).

174. Golay, X., Hendrikse, J. & Lim, T. C. Perfusion imaging using arterial spin labeling. *Top. Magn. Reson. Imaging* **15**, 10–27 (2004).

175. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).

176. MacInnes, J. J., Dickerson, K. C., Chen, N. K. & Adcock, R. A. Cognitive neurostimulation: learning to volitionally sustain ventral tegmental area activation. *Neuron* **89**, 1331–1342 (2016).

177. Bauer, C. C. C. et al. Real-time fMRI neurofeedback reduces auditory hallucinations and modulates resting state connectivity of involved brain regions: part 2: default mode network — preliminary evidence. *Psychiatry Res.* **284**, 112770 (2020).

178. Tarakad, A. & Jankovic, J. Anosmia and ageusia in Parkinson's disease. *Int. Rev. Neurobiol.* **133**, 541–556 (2017).

179. Williams, Z. M., Bush, G., Rauch, S. L., Cosgrove, G. R. & Eskandar, E. N. Human anterior cingulate neurons and the integration of monetary reward with motor responses. *Nat. Neurosci.* **7**, 1370–1375 (2004).

180. Sheth, S. A. et al. Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. *Nature* **488**, 218–221 (2012).

181. Shine, J. M. et al. Distinct patterns of temporal and directional connectivity among intrinsic networks in the human brain. *J. Neurosci.* **37**, 9667–9674 (2017).

182. Rafiei, F., Safrin, M., Wokke, M. E., Lau, H. & Rahnev, D. Transcranial magnetic stimulation alters multivoxel patterns in the absence of overall activity changes. *Hum. Brain Mapp.* **42**, 3804–3820 (2021).

183. Buzsáki, G., Anastassiou, C. A. & Koch, C. The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* **13**, 407–420 (2012).

184. Granger, C. W. J. & Newbold, P. *Forecasting Economic Time Series* (Academic, 1977).

185. Friston, K. J., Harrison, L. & Penny, W. Dynamic causal modelling. *Neuroimage* **19**, 1273–1302 (2003).

186. Barnett, L., Barrett, A. B. & Seth, A. K. Misunderstandings regarding the application of Granger causality in neuroscience. *Proc. Natl Acad. Sci. USA* **115**, E6676–E6677 (2018).

187. Friston, K. Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLoS Biol.* **7**, e33 (2009).

188. Webb, J. T., Ferguson, M. A., Nielsen, J. A. & Anderson, J. S. BOLD granger causality reflects vascular anatomy. *PLoS ONE* **8**, e84279 (2013).

189. Granger, C. W. Time series analysis, cointegration, and applications. *Am. Econ. Rev.* **94**, 421–425 (2004).

190. Mehler, D. & Kording, K. The lure of causal statements: rampant mis-inference of causality in estimated connectivity. Preprint at https://arxiv.org/abs/1812.03363 (2018).

Author contributions
S.H.S and M.D.F. wrote the manuscript. The authors all researched data for the article, provided substantial contributions to discussion of its content, and reviewed and edited the manuscript before submission.

Competing interests
S.H.S. serves as a consultant for Kaizen Brain Center, Acacia Mental Health and Magnus Medical, and owns stock in Brainsway Inc. (publicly traded) and Magnus Medical (not publicly traded). S.H.S. and M.D.F. have jointly received investigator-initiated research support from Neuronetics Inc. None of these organizations were involved in the present work. S.H.S. and M.D.F. each own independent intellectual property on the use of brain network mapping to target neuromodulation. The present work did not utilize any of this intellectual property. The authors report no other conflicts of interest related to the present work.

Peer review information
Nature Reviews Neuroscience thanks G. Hartwigsen, A. Kiyonaga and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.