Trends in Neurosciences

CellPress OPEN ACCESS

Opinion

Bridging the big (data) gap: levels of control in small- and large-scale cognitive neuroscience research

Roni Tibon [1,2,*] Linda Geerligs,[3] and Karen Campbell[4]

Recently, cognitive neuroscience has experienced unprecedented growth in the availability of large-scale datasets. These developments hold great methodological and theoretical promise: they allow increased statistical power, the use of nonparametric and generative models, the examination of individual differences, and more. Nevertheless, unlike most 'traditional' cognitive neuroscience research, which uses controlled experimental designs, large-scale projects often collect neuroimaging data not directly related to a particular task (e.g., resting state). This creates a gap between small- and large-scale studies that is not solely due to differences in sample size. Measures obtained with large-scale studies might tap into different neurocognitive mechanisms and thus show little overlap with the mechanisms probed by small-scale studies. In this opinion article, we aim to address this gap and its potential implications for the interpretation of research findings in cognitive neuroscience.

Highlights

Studies in the field of cognitive neuroscience vary along two different axes: the first refers to their 'levels of control' (e.g., controlled task designs vs. resting-state paradigms) and the second refers to their sample size.

Although theoretically these two axes can be orthogonal, in practice they often correlate: controlled designs generally dominate the field for small-scale studies, whereas uncontrolled designs are prevalent among large-scale studies.

This results in qualitative differences between small- and large-scale studies, which are not due to their size *per se*, and poses challenges to the interpretation of the results.

Two methodological axes of cognitive neuroscience research

Studies in the field of cognitive neuroscience vary along two different axes. The first axis, termed here 'levels of control', refers to the nature of the designs that are employed: in some studies, the cognitive processes are driven by external input provided by the researcher (i.e., particular tasks or stimuli). Tight control over the experimental input allows the researcher to track the specific cognitive processes elicited by the task. At the other end of this continuum are studies in which cognitive processes are generated internally (i.e., without any external stimulus or task). With no external input, and without the ability to infer what participants are thinking, such studies entail low levels of control.

The second axis refers to the size of the study: traditionally, studies in the field of cognitive neuroscience have relied on data collected from relatively few participants (often fewer than 20 per group, although see the later discussion of current trends). However, in recent years, large-scale studies, for which data are collected from hundreds or even thousands of participants, are becoming increasingly popular.

Although theoretically these two axes can be orthogonal, in practice they often correlate. In this opinion article, we investigate these two axes in the context of recent advances in the field of cognitive neuroscience and demonstrate how these qualitative differences between small- and large-scale studies might pose considerable challenges to the interpretation of research results.

Levels of control in cognitive neuroscience

Studies in the field of cognitive neuroscience employ different levels of control. The traditional approach in the field is that tightly controlled experimental manipulations can be applied to discover

[1]MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK
[2]School of Psychology, University of Nottingham, Nottingham, UK
[3]Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, The Netherlands
[4]Department of Psychology, Brock University, St. Catharines, Ontario, Canada

*Correspondence: roni.tibon@nottingham.ac.uk (R. Tibon).

interpretable relationships between the brain and behavior. This notion relies on the assumption that carefully designed experimental tasks can be used to individuate cognitive processes, which can then be linked to specific neural mechanisms. For example, a central debate in the cognitive neuroscience of episodic memory revolved around the role of the hippocampus in episodic retrieval: whether it selectively supports recollection (i.e., the retrieval of qualitative information about a specific study episode), or otherwise supports the retrieval of 'strong' memories, regardless of whether their retrieval is achieved via recollection or familiarity-based judgments of prior occurrence [1]. By combining neuroimaging measures with a range of tightly controlled memory tasks (including source memory, context memory, and continuous recognition tasks), studies lent support to the notion that the hippocampus tracks the amount of episodic information that is retrieved, thereby selectively supporting recollection, rather than the strength of an undifferentiated memory signal [2]. Other examples can be found in the fields of face perception (the face specificity vs. individuation/expertise debate [3–7]), visual working memory (representations maintained in frontoparietal vs. occipital regions [8–10]), and syntactic processing (a specific role in syntactic processing vs. a general role in cognition for the left inferior frontal gyrus [11]). Over the years, this combination of neuroimaging measures and tightly controlled experimental designs allowed cognitive neuroscientists to make immense progress in characterizing and understanding the relations between cognitive and neural mechanisms via multiple imaging modalities (Box 1) and across multiple cognitive domains.

At the other end of the levels-of-control continuum, resting-state studies emerged as one form of uncontrolled design, often used as an alternative approach to map the relationship between trait-like brain networks and cognitive functions. These studies build on the important discovery that consistent functional resting-state networks are observed in the absence of any stimulus or task [12–14]. In a typical resting-state study, neural data are acquired while participants are not exposed to any controlled external input (i.e., specific task or stimuli). In some cases, additional data are collected (e.g., data from behavioral tasks performed outside the scanner) and linked to neural patterns that are observed at rest. This approach has been widely applied to the study of various populations and clinical applications [15]. In addition, commonalities between

Box 1. Functional neuroimaging modalities in cognitive neuroscience

A variety of neuroimaging modalities, including fMRI, electroencephalography (EEG), MEG, positron emission tomography (PET)/computed tomography (CT), and functional near-IR spectroscopy (fNIRS), are used in the field of cognitive neuroscience, often providing different perspectives or complementary evidence that relies on the unique strengths offered by each technique. Nevertheless, in large-scale projects, there is a clear dominance of fMRI. In a recent review of open-access neuroimaging datasets [26], of the 61 large-scale projects ($n > 100$) involving human participants, 37 included fMRI data but only nine included EEG and/or MEG data (we are aware of only two additional large-scale projects involving M/EEG data, which were not included in that review [48,49]).

Why some modalities are being relatively overlooked in large-scale projects is not entirely clear. Possible reasons for the popularity of fMRI in large-scale projects include the straightforward combination with structural brain metrics and the greater spatial resolution of fMRI. Nevertheless, given that EEG is widely available and frequently used in small-scale studies, while also entailing relatively low costs, one might expect EEG to be at least as popular as fMRI for large-scale projects. Thus, another reason for the popularity of fMRI in large-scale projects may stem from the fact that resting-state networks were first established with fMRI [12–14], and resting-state data are relatively easy to collect on a grand scale (see main text). Although resting-state networks have also been observed using MEG/EEG [50–54], the typical way of analyzing such data remains task focused, and there is less consensus regarding resting-state connectivity metrics [55]. Nevertheless, the investigation of the temporal dynamics of resting-state networks might greatly benefit from the superior temporal resolution of these modalities.

Despite this dominance of fMRI in large-scale projects, it might not be the optimal modality for every purpose. It is important to consider, we would argue, a range of methods, and rather than seeing fMRI as a 'default' modality, to carefully select the methods that seem best suited to the specific research question(s) at hand.

resting-state networks and task activation networks have been observed [16,17], suggesting that both are largely shaped by a shared and stable intrinsic network architecture [18]. It is important to note, however, that despite the prevalence of resting-state designs, especially in large-scale studies, resting-state data are not unique in how they can be analyzed, and methods that measure functional connectivity can be similarly applied to task data [19,20]. In addition to tightly controlled experimental designs and resting-state designs, other designs are becoming increasingly popular (Box 2), providing additional opportunities for the investigation of human neurocognition.

Small- versus large-scale studies

The number of participants in much of the cognitive neuroscience literature tends to be low. A recent study [21] scrutinizing sample sizes in neuroimaging revealed that, between 1990 and 2021, the median sample size in highly cited experimental functional magnetic resonance imaging (fMRI) studies was ~12 per group. The study also reported a consistent growth in the median sample size over the years, up to 24 in 2018. This growth can be attributed to the impact of the reproducibility crisis (Box 3), which stressed the necessity of larger samples (or more observations within participants [22–25]), as well as to the growing availability and ease of use of experimental and neuroimaging resources. Despite this increase, however, the sample size for most experimental neuroimaging studies (i.e., those that employ controlled designs) rarely exceeds a few dozen participants.

Despite the fact that most experimental studies still rely on relatively low numbers of participants, the recent movement towards large-scale ('big data') initiatives that is evident in many fields is also present in the field of cognitive neuroscience. Several large-scale studies (recently reviewed

Box 2. Naturalistic designs: beyond controlled tasks and resting-state data

In the field of cognitive neuroscience, the recent decade has been characterized by increasing methodological versatility. This has provided opportunities for innovative study designs that are different from the more traditional ones. In particular, designs that incorporate naturalistic stimuli or behaviors are gaining popularity. For example, in movie-viewing paradigms, participants watch a movie without performing any particular task (other than following a general instruction to 'watch the movie') while neural measures are collected [35,56–59]. Similarly, in ongoing narrative-processing paradigms, participants might listen to a story, a conversation, or a lecture, again provided with some general or minimal instructions [59,60]. Such tasks comprise a continuous input stream, providing a more ecological model for how information is obtained and processed in 'real-world' situations. Arguably, however, the naturalistic characteristics of such tasks are constrained by the fact that, unlike our real-world experiences, in which we act and interact with the stimuli that surround us, the range of behaviors afforded by these tasks remains limited. To account for these limitations, recent research has incorporated interactive behaviors, such as continuous speech or navigation in a virtual-reality environment, to better approach naturalistic cognition [61–63]. Although the production of naturalistic behaviors within the confines of an fMRI or MEG scanner is undoubtedly challenging, such research benefits from recent methodological developments that allow better mobility of neuroimaging techniques, such as mobile EEG devices and wearable MEG systems [64–67].

One of the initial motivations to use such 'naturalistic' designs was rooted in the long-standing critique that because of the isolated and artificial nature of tightly controlled tasks, the conclusions they generate often fail to generalize outside the laboratory into real-life contexts [68–73]. Moreover, although naturalistic paradigms do not have the experimental efficiency of targeted manipulations [71], they offer better understanding of the relative contributions of different domains to real-world phenomena [74]. Importantly, these designs also overcome some of the limitations of uncontrolled resting-state designs [19], for example, by maintaining the complex, multidimensional, and ongoing nature of sensory input that the brain has evolved to capitalize on to guide behavior [69]. Naturalistic stimuli have also been shown to enhance meaningful individual differences in functional connectivity [75] and provide more direct links between neural and cognitive mechanisms than resting-state studies (e.g., by using annotated stimulus features and narrative structure [20,57,76–78]). In the context of large-scale studies, the great advantage of naturalistic paradigms is the richness of the data they produce. This richness allows these data to be used to answer a wider range of research questions than typical task-controlled or resting-state studies. For example, a spoken narrative can be used to investigate questions ranging from the domain of word meaning or syntactic processing to the perception of emotions and social cognition [69,71,79]. At the practical level (i.e., 'ease of use'), these designs are comparable with resting-state designs and therefore are relatively easy to implement in large-scale studies, and have been shown to reduce head motion in the scanner relative to rest [80]. When these are complemented by controlled tasks, datasets will also allow more opportunities to map abstract laboratory-based tasks onto more real-world cognitive processing.

Box 3. The reproducibility crisis and samples that are too small

The reproducibility crisis in science, and in psychological sciences in particular, refers to the repeated observation that research findings often fail to replicate [81–83]. This failure has been attributed to multiple reasons, including poorly conducted procedures, flawed career incentive structures, and biases in the publication system [84–87]. Among these issues, the most frequently identified cause is lack of statistical power due to samples that are too small [40,85]. In the null hypothesis significance testing framework, statistical power is defined as the probability of rejecting the null hypothesis (i.e., obtaining a statistically significant test result) given that the alternative hypothesis is true.

As discussed in previous articles [21,40,85], the problem with underpowered studies is threefold. First, inherent to the aforementioned definition of power, in underpowered studies, the chance of discovering effects that are genuinely true is low. That is, low-powered studies produce more false negatives than high-powered studies. Second, low power reduces the positive predictive value (PPV) of the study, defined as the probability that a 'positive' research finding reflects a true effect (i.e., the finding is a true positive). For example, a PPV of 80% means that 80% of a study's claims for discoveries will be correct. The link between this probability and the power is indicated by the formula

$$\mathrm{PPV} = [(1 - \beta) \times R]/[(1 - \beta) \times R + \alpha], \quad [\mathrm{I}]$$

where $(1 - \beta)$ is the power, α is the type I error, and R is the pre-study odds. With the other components of the formula (α and R) held constant, the greater the power the greater the PPV.

Third, underpowered studies are likely to yield an exaggerated estimate of the magnitude of the effect when a true effect is discovered. To clarify this point, suppose, for example, that a medium-sized effect (e.g., $d = 0.5$) truly exists. Due to sampling variations and random errors, studies addressing this effect will yield effect sizes that fluctuate around the actual effect size (e.g., between 0.4 and 0.6). However, because of the small sample size and a predefined threshold of statistical significance (e.g., $P < 0.05$), only studies that produced a large effect size (due to these random fluctuations) would reach statistical significance.

Underpowered studies (due to small sample sizes) are common in the field of cognitive neuroscience [21]. This lack of power is partially attributed to the major costs associated with the collection of neuroimaging data, but also to other non-financial factors such as ambiguity around how to define effect size or power in fMRI. Although in recent years sample sizes have been increasing in neuroimaging research [21], the more fundamental transformations that could bring the replication crisis to an end remain to materialize.

in [26]) that involve the acquisition of neuroimaging (and other) data from a large number of participants (at least a few hundred) have been launched in recent years. For example, the WU/Minn Project [27] [part of the Human Connectome Project (HCP)] comprises data collected from ~1200 participants (twin siblings). Functional neuroimaging measures in the HCP include R-fMRI as well as seven well-established (although arguably basic) fMRI tasks tapping into multiple cognitive domains (working memory, reward processing, motor processing, language, social cognition, relational processing, and emotion processing). For working memory, for example, a common variant of the widely used *n*-back task was employed, with high ('2-back') versus low ('0-back') working memory load. Another large-scale project is the UK Biobank [28,29]. This biomedical database and research resource contains in-depth genetic and health information from half a million UK participants. Neuroimaging measures (from n = ~5000 participants) include R-fMRI data and task fMRI data from a well-established faces/shapes 'emotion' task [30,31] in which participants are shown a visual stimulus at the top of the screen (either a shape or an emotionally negative face) and are asked to decide which of two shapes/faces presented at the bottom of the screen match that stimulus. A final example is the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) project, which uses epidemiological, cognitive, and neuroimaging data to understand how individuals can best retain cognitive abilities into old age [32,33]. This dataset includes multimodal [fMRI and magnetoencephalography (MEG)] resting-state and sensory-motor task data, as well as fMRI movie-watching data, from n = ~700 participants aged 18–88 years. In addition, it includes MEG and fMRI measures from a variety of tasks tapping into multiple cognitive domains (including emotion regulation, emotional memory,

fluid intelligence, picture naming, response selection and inhibition, sentence comprehension, and visual short-term memory), but only for a subset of the participants (n = ~140 per task).

Such projects hold great methodological and theoretical promise: they allow increased statistical power and the examination of individual differences. Large datasets also enable the use of models with higher complexity, such as nonparametric and generative models that aim to generate new data consistent with the original observations [34]. However, as we discuss later, the transition from small- to large-scale studies is not always straightforward. In particular, when upscaling, the level of control in the task designs that are implemented is often sacrificed due to practical considerations.

Beyond sample size: qualitative differences between small- and large-scale studies

The examples discussed earlier suggest that the two axes that can be used to characterize studies in the field of cognitive neuroscience – level of control and sample size – are often correlated. Tightly controlled designs dominate the field for small-scale studies. These often involve innovative tasks that, to provide novel insights into cognitive (sub)processes, are typically fine-tuned to probe highly specific aspects of cognition. These tasks are sometimes lengthy and might include several stages (e.g., encoding and retrieval for memory tasks) and/or high levels of complexity (e.g., multiple experimental conditions). By contrast, many of the large-scale projects are dominated by uncontrolled resting-state designs coupled with relatively simple and well-validated controlled tasks, typically tapping into coarse or well-characterized cognitive processes. Therefore, in most cases, the transition from small-scale to large-scale studies does not simply mean 'more data'. Instead, large-scale studies tend to collect data using designs that are qualitatively different from those that are typically used for small-scale studies.

The discrepancy between small- and large-scale studies does not seem to result from theoretical considerations but more likely reflects practical ones. Namely, although controlled task designs, especially when combined with neuroimaging measures, allow some flexibility at the analytical level (e.g., employing univariate and multivariate analyses to investigate variations in task-induced processes and representations), they often aim to examine a limited number of cognitive (sub)processes. On its own, this narrow focus might not be a limitation, and indeed reflects an inherent property of these designs – that is, using well-matched tasks to investigate highly specific cognitive processes. This 'feature', however, turns into a limitation in the execution of large-scale projects: to maximize their potential contribution, and given the time, effort, and resources they require, an important characteristic of large-scale projects is that they are able to serve multiple purposes. Controlled designs, however, are unlikely to be informative beyond the specific cognitive processes that they were designed to probe. By contrast, uncontrolled designs often aim to tap into myriad cognitive processes and can therefore be used more flexibly and to allow more versatility, thereby maximizing the potential 'return' of a large-scale project. Another advantage of uncontrolled designs, which is likely to constitute a crucial consideration for large-scale projects, is that with these design data are relatively easy to obtain, manage, and share [19,35]. This stands in contrast to the often nuanced and complex nature of controlled designs, which can be difficult to interpret for non-experts in a particular cognitive field.

To a large extent, these practical considerations have dictated the scope and type of tasks used in large-scale studies. In some of these studies, controlled designs are also included, but these tend to be short, simple, and designed to achieve goals markedly different from those commonly encountered in small-scale studies. For instance, the fMRI tasks included in the HCP were included '(i) to help identify as many 'nodes' as possible that can guide, validate, and interpret

the results of the connectivity analyses that will be conducted on resting-state fMRI (R-fMRI), resting-state MEG (R-MEG), and diffusion data; (ii) to allow a comparison of network connectivity in a task context to connectivity results generated using R-fMRI; and (iii) to relate signatures of activation magnitude or location in key network nodes to individual differences in performance, psychometric measures, or other phenotypic traits' ([31], see p. 170). These goals are quite different from those of a typical small-scale study, which tend to be aimed at determining how specific cognitive processes are instantiated in the brain or at testing specific psychological theories [36].

Study goals are important because they dictate the types of tasks used. For the HCP, the tasks were designed to act as 'functional localizers' for a diverse range of cognitive functions (e.g., working memory, motor control, emotion processing) and were selected for their reliability and efficiency (most tasks were ~3 min in length and used a block design). While these tasks were well suited to the goals of the HCP, they are unlikely to be particularly informative about the specific cognitive functions performed by individual regions in a given network. For instance, the language processing task contrasts four blocks of listening to a story with four blocks of solving mathematics problems. Such a design does not allow isolation of the different component processes involved in language (i.e., syntax, semantics, and pragmatics), for which a more nuanced design is required [37].

Further, the utility of these tasks in predicting individual differences in cognitive performance depends on whether the processes that differ across individuals are taxed by a given task. For instance, age differences in motor control may not be observed during a simple task that requires one movement at a time (i.e., tap right or left fingers, squeeze left or right toes) but may emerge only when different movements conflict or when a prepotent response must be inhibited [38]. Overly simple tasks can miss out on individual differences in the ability to reconfigure network interactions to accomplish more challenging task goals [39] and may also be unsuited to identify the neural underpinnings of specific cognitive disorders and developmental changes.

What are the insights that large-scale studies do offer?

The high specificity and novel insights into cognitive processes offered by controlled designs are not readily obtained with current large-scale studies. This is not to say that these studies do not offer novel perspectives, but they differ in the type of knowledge they contribute. Namely, while current small-scale and well-controlled studies are useful for identifying specific neurocognitive processes, large-scale studies can help to, for instance, further characterize processes that are already known (by using a combination of uncontrolled and well-established controlled designs) and identify individual differences therein.

The current landscape, in which small-scale studies tend to rely on controlled designs whereas large-scale studies tend to rely on uncontrolled designs, poses challenges to the interpretation of the results. When commonalities between small- and large-scale studies emerge, they are often informative and can provide complementary perspectives on cognitive processes. For example, suppose that a small-scale study revealed that a certain memory task involves greater connectivity within some brain network for remembered versus forgotten items. Now suppose that in a large-scale study, resting-state connectivity within the same network was related to task performance in a memory task performed outside the scanner. Taken together, these findings provide compelling evidence that this network supports memory. However, when differences emerge between small- and large-scale studies, their interpretation is often hindered by the qualitative differences in the designs. Suppose, again, that in a large-scale study, the aforementioned brain network was not related to memory performance. Is that because this network is

activated only during the task? Is it specific to the subprocesses probed by the memory task? Is it because the projects used different cohorts or is the observed effect spurious due to the small sample of the small-scale study [40]? These alternatives are difficult to disentangle.

To elevate current understanding of cognitive functions to the next level, large-scale studies would benefit from the inclusion of more fine-grained tasks. With that, due to the substantial resources required for the collection of large-scale datasets, it makes little sense to include tasks that are not carefully validated in terms of their test–retest reliability, their construct validity, and their measurements. Reliable measurements are a prerequisite for studies of inter-individual variability. Does this imply that one should only use well-validated coarse tasks that are unlikely to provide new insights? Should large-scale studies abandon novel fine-grained tasks that can expand our knowledge but might not be as reliable? We would argue that with additional planning and creativity, middle-ground solutions can be formed. For example, data collection can be done in phases, where data from novel tasks collected in each phase are analyzed prior to the execution of the next phase. This way, tasks can be replaced if found to be unsuitable. Such solutions might require additional resources in the short term (e.g., additional contingency planning, allowing more time for data collection, etc.), but overall they can reduce the time and costs involved in scanning hundreds/thousands of individuals on a suboptimal task.

We would argue that the few large-scale studies that do include a variety of neurocognitive measures, such as any combination of resting-state, coarse and fine tasks, and other (e.g., naturalistic) designs, already demonstrate the utility of this approach and illustrate how, when integrated, these measures can be used to validate previous results and to lead to new insights in the field of cognitive neuroscience. In one study, which used data from Cam-CAN [32,33], functional connectivity during three mental states (resting-state, movie watching, and a sensorimotor task) was measured within the same participants [41]. In accord with previous findings, the authors showed commonalities between connectivity patterns across states. Importantly, however, the study further revealed divergence between the states, suggesting that individual differences in functional connectivity are driven not only by trait-like aspects but also by state-dependent ones. Another study [42] analyzed fMRI data from seven tasks that tap into different domains, to examine whole-brain task-related modulation of functional connectivity. By leveraging large-scale task fMRI data from the HCP [31] and the University of California Los Angeles (UCLA) Consortium for Neuropsychiatric Phenomics [43], the study showed that task-related modulations of connectivity patterns are not restricted by regional task activations, thus reinforcing the importance of studying whole-brain task connectomes. In another example [44], task and resting-state HCP data were used to validate previous findings showing correspondence between brain networks that are activated during task and rest, while also generating new results regarding the engagement of different brain networks in particular tasks. In a final example [45], independent component analysis (ICA) was applied to task data from Cam-CAN to derive networks that are shared and unique across two tasks that typically decline with age (fluid intelligence and object naming) and one that is typically preserved (sentence comprehension). It was found that the same networks that showed an age-related decrease in activity and connectivity in the declining tasks did not show an age difference in the maintained task, suggesting that age differences in network function depend on the task context. These examples demonstrate how the inclusion of multiple neurocognitive measures in a single large-scale initiative can offer novel insights and enhanced interpretability that go beyond what can be obtained via 'local' small-scale studies or via large-scale studies that do not include controlled designs. In particular, we believe that the inclusion of novel task designs in large-scale studies is essential to provide novel insights regarding specific processes within cognitive domains and would therefore play a key role in driving the field forward.

Concluding remarks

During the past decade, big data has become one of the hottest buzzwords in many industries and research areas, and the field of cognitive neuroscience is no exception. Despite this growing popularity of large-scale datasets, as researchers we should upscale with caution (see Box 4 for suggestions and see Outstanding questions for additional considerations). In a reality of limited resources, progress might be achieved by our ability to 'go bigger', but might also depend on other factors. Large-scale projects require many resources and their data collection and analysis is slower and more tedious. In addition, their data are often analyzed many times by different research teams, which enhances the risk of increasing false positives and for our research questions being limited by the data that are available [26,46,47] (see Outstanding questions). Another challenge is that methodological progress may render some of the technology currently used in large-scale studies less relevant in the future, complicating the comparison of findings over time. Importantly, as discussed earlier, current small- and large-scale studies can make differential contributions to the field of cognitive neuroscience, owing to, among other factors, variations in the nature of the designs that they typically use.

In a way, this might be unsurprising: small-scale studies are usually designed within the field of cognitive neuroscience by cognitive neuroscientists and are tuned to provide new insights on cognition and its neural substrates. By contrast, large-scale studies are usually designed by multidisciplinary teams, which include cognitive neuroscientists but also researchers from other domains, and aim to tackle a variety of topics, with cognition often included as a secondary goal if at all. As well justified as these differences between small- and large-scale studies might be, the important implication of what one can or cannot learn from a particular dataset should be taken into account.

Box 4. Super-size me: how to upscale?

Large-scale studies often employ uncontrolled task designs. Nevertheless, to bridge the gap between small- and large-scale studies, alternative designs should be considered. One option is to collect large-scale data for controlled designs. The main advantage of this route is that small and large cohorts are then directly comparable. However, because controlled designs are both costly (i.e., tasks are often lengthy and complex) and inflexible (i.e., targeting highly specific cognitive processes), for large-scale projects, the limited efficiency of this route often outweighs its potential advantages. Another alternative is to collect large-scale data for 'known' cognitive processes (i.e., processes that were already probed via 'classic' cognitive tasks) using naturalistic designs (Box 2). For example, continuous narrative processing can be used to tap into a variety of domains, such as auditory processing, language, memory, emotion, etc. Stimulus features (e.g., emotional themes, specific auditory properties) can be labelled and contribute to event-based analyses [57,61,88–92]. Additional tasks performed outside the scanner (e.g., a memory task about the content of the stimuli) can also be used to facilitate interpretation. Thus, although novel insights into subtle subprocesses (e.g., those offered by task designs) are hard to obtain using this approach, it allows understanding of how (relatively coarse) processes would operate while also allowing more flexibility and easier implementation than controlled experimental designs. A similar route entails the collection of large-scale data using naturalistic paradigms, together with controlled data that are also collected in the same large-scale study. Arguably, this route represents an enhanced solution in terms of interpretability, but it incurs higher costs. To overcome this problem, controlled task data can be collected for different subsets of participants (i.e., with each participant taking part in some of the controlled tasks included in the study while all of the participants contribute to naturalistic data). The advantage of this solution is that it maintains some overlap between the cohorts for the different designs and therefore allows direct comparison/links between them (see Cam-CAN [32,33] for a similar implementation).

Alternative routes for upscaling also exist, and different routes might be useful for different purposes. For example, a cohort study aiming to investigate aging effects might benefit more from employing uncontrolled and naturalistic designs as well as well-established tasks, whereas a project aiming to map the neurocognitive mechanisms underlying a proposed taxonomy of memory processes might benefit from a combination of novel and well-established tightly controlled tasks, together with naturalistic designs. In some cases, a large *n* is necessary to meet the study's goals (e.g., when individual differences are of crucial interest), whereas in other cases, the study might benefit more from dense sampling (i.e., testing a relatively low number of individuals on many experimental tasks/conditions) [22–25]. Importantly, when designing and analyzing large-scale projects, these important variations and their outcomes should be kept in mind.

Outstanding questions

How do we determine the ideal transition route (or design) for a planned large-scale study? Given the large investment these studies require, and their usefulness to the wider scientific community (as open datasets), should the planning of these studies be more community driven – for instance, by soliciting potential cognitive tasks from various research groups and putting them to a vote via a scientific society such as the Organization for Human Brain Mapping?

In a field with fast-evolving technology (i.e., new equipment, new scanning sequences, etc.), how can we best avoid using methods that may eventually become obsolete to maximize the usefulness of data collected in large-scale studies for years to come?

It would be beneficial for the field to decide on some shared defaults or guidelines for large-scale studies. First, what are default or common stimuli/tasks that could be used instead of (or in addition to) the resting state? Second, what should be the default infrastructures and standards for the sharing and dissemination of behavioral and neuroimaging data?

The benefit of large-scale projects is that the data can be used for many different studies to answer a wide range of research questions. However, previous work has shown that data reuse can increase the number of false positives, since researchers do not correct for the increased number of multiple comparisons across studies. To what extent do the benefits of well-powered studies outweigh these potential risks of systematic bias across many studies?

With this in mind, we believe that, in addition to uncontrolled resting-state data, more elaborate task-based measures may offer several advantages. Science relies on our ability to generate and test new hypotheses. As knowledge in the complementary fields of cognition, emotion, and developmental psychology continues to evolve, our paradigms and measures in cognitive neuroscience must also do so. Ultimately, we want to understand how brain function drives cognition in the real-world settings it evolved to navigate; well-powered studies with a mix of controlled, uncontrolled, and naturalistic tasks bring us closer to realizing this aim.

Acknowledgments

R.T. was supported by a British Academy Postdoctoral Fellowship (PF170046: SUAI/028 RG94188). L.G. was supported by a Vidi grant from the Dutch Research Council (NWO, VI.Vidi.201.150). K.C. was supported by the Canada Research Chairs program.

Declaration of interests

The authors declare no interests.

References

1. Moscovitch, M. *et al.* (2016) Episodic memory and beyond: the hippocampus and neocortex in transformation. *Annu. Rev. Psychol.* 67, 105–134
2. Rugg, M.D. *et al.* (2012) Item memory, context memory and the hippocampus: fMRI evidence. *Neuropsychologia* 50, 3070–3079
3. Kanwisher, N. (2000) Domain specificity in face perception. *Nat. Neurosci.* 3, 759–763
4. Kanwisher, N. *et al.* (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311
5. Grill-Spector, K. *et al.* (2004) The fusiform face area subserves face perception, not generic within-category identification. *Nat. Neurosci.* 7, 555–562
6. Tarr, M.J. and Gauthier, I. (2000) FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat. Neurosci.* 3, 764–769
7. Gauthier, I. *et al.* (1999) Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nat. Neurosci.* 2, 568–573
8. Teng, C. and Postle, B.R. (2021) Understanding occipital and parietal contributions to visual working memory: commentary on Xu (2020). *Vis. Cogn.* 29, 401–408
9. Scimeca, J.M. *et al.* (2018) Reaffirming the sensory recruitment account of working memory. *Trends Cogn. Sci.* 22, 190–192
10. Gayet, S. *et al.* (2018) Visual working memory storage recruits sensory processing areas. *Trends Cogn. Sci.* 22, 189–190
11. Tyler, L.K. *et al.* (2011) Left inferior frontal cortex and syntax: function, structure and behaviour in patients with left hemisphere damage. *Brain* 134, 415–431
12. Damoiseaux, J.S. *et al.* (2006) Consistent resting-state networks across healthy subjects. *Proc. Natl. Acad. Sci. U. S. A.* 103, 13848–13853
13. Raichle, M.E. *et al.* (2001) A default mode of brain function. *Proc. Natl. Acad. Sci. U. S. A.* 98, 676–682
14. Greicius, M.D. *et al.* (2003) Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* 100, 253–258
15. Arbabshirani, M. *et al.* (2013) Classification of schizophrenia patients based on resting-state functional network connectivity. *Front. Neurosci.* 7, 133
16. Smith, S.M. *et al.* (2009) Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci. U. S. A.* 106, 13040–13045
17. Tavor, I. *et al.* (2016) Task-free MRI predicts individual differences in brain activity during task performance. *Science* 352, 216–220
18. Cole, M.W. *et al.* (2014) Intrinsic and task-evoked network architectures of the human brain. *Neuron* 83, 238–251
19. Campbell, K.L. and Schacter, D.L. (2017) Ageing and the resting state: is cognition obsolete? *Lang. Cogn. Neurosci.* 32, 661–668
20. Finn, E.S. (2021) Is it time to put rest to rest? *Trends Cogn. Sci.* 25, 1021–1032
21. Szucs, D. and Ioannidis, J.P.A. (2020) Sample size evolution in neuroimaging research: an evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *NeuroImage* 221, 117164
22. Chen, G. *et al.* (2022) Hyperbolic trade-off: the importance of balancing trial and subject sample sizes in neuroimaging. *NeuroImage* 247, 118786
23. Nee, D.E. (2019) fMRI replicability depends upon sufficient individual-level data. *Commun. Biol.* 2, 130
24. Naselaris, T. *et al.* (2021) Extensive sampling for complete models of individual brains. *Curr. Opin. Behav. Sci.* 40, 45–51
25. Pinho, A.L. *et al.* (2018) Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping. *Sci. Data* 5, 180105
26. Madan, C.R. (2021) Scan once, analyse many: using large open-access neuroimaging datasets to understand the brain. *Neuroinformatics* Published online May 11, 2021. https://doi.org/10.1007/s12021-021-09519-6
27. Van Essen, D.C. *et al.* (2013) The WU-Minn Human Connectome Project: an overview. *NeuroImage* 80, 62–79
28. Bycroft, C. *et al.* (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209
29. Miller, K.L. *et al.* (2016) Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* 19, 1523–1536
30. Hariri, A.R. *et al.* (2002) The amygdala response to emotional stimuli: a comparison of faces and scenes. *NeuroImage* 17, 317–323
31. Barch, D.M. *et al.* (2013) Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage* 80, 169–189
32. Shafto, M.A. *et al.* (2014) The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC Neurol.* 14, 204
33. Taylor, J.R. *et al.* (2017) The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage* 144, 262–269
34. Bzdok, D. and Yeo, B.T.T. (2017) Inference in the age of big data: future perspectives on neuroscience. *NeuroImage* 155, 549–564
35. Finn, E.S. *et al.* (2020) Idiosynchrony: from shared responses to individual differences during naturalistic neuroimaging. *NeuroImage* 215, 116828
36. Henson, R. (2005) What can functional neuroimaging tell the experimental psychologist? *Q. J. Exp. Psychol. A* 58, 193–233
37. Campbell, K.L. and Tyler, L.K. (2018) Language-related domain-specific and domain-general systems in the human brain. *Curr. Opin. Behav. Sci.* 21, 132–137

38. Sebastian, A. *et al.* (2013) Differential effects of age on subcomponents of response inhibition. *Neurobiol. Aging* 34, 2183–2193
39. Davis, S.W. *et al.* (2017) Resting-state networks do not determine cognitive function networks: a commentary on Campbell and Schacter (2016). *Lang. Cogn. Neurosci.* 32, 669–673
40. Button, K.S. *et al.* (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376
41. Geerligs, L. *et al.* (2015) State and trait components of functional connectivity: individual differences vary with mental state. *J. Neurosci.* 35, 13949–13961
42. Di, X. and Biswal, B.B. (2019) Toward task connectomics: examining whole-brain task modulated connectivity in different task domains. *Cereb. Cortex* 29, 1572–1583
43. Poldrack, R.A. *et al.* (2016) A phenome-wide examination of neural and cognitive function. *Sci. Data* 3, 160110
44. Nickerson, L.D. (2018) Replication of resting state–task network correspondence and novel findings on brain network activation during task fMRI in the Human Connectome Project study. *Sci. Rep.* 8, 17543
45. Samu, D. *et al.* (2017) Preserved cognitive functions with age are determined by domain-dependent shifts in network responsivity. *Nat. Commun.* 8, 14743
46. Kievit, R.A. *et al.* (2022) Using large, publicly available data sets to study adolescent development: opportunities and challenges. *Curr. Opin. Psychol.* 44, 303–308
47. Thompson, W.H. *et al.* (2020) Meta-research: dataset decay and the problem of sequential analyses on open datasets. *eLife* 9, e53498
48. Harati, A. *et al.* (2014) The TUH EEG CORPUS: a big data resource for automated EEG interpretation. In *2014 Signal Processing in Medicine and Biology Symposium (SPMB)*, pp. 1–5, IEEE
49. Vaghari, D. *et al.* (2021) A multi-site, multi-participant magnetoencephalography resting-state dataset to study dementia: the BioFIND dataset. *medRxiv* Published online May 19, 2021. https://doi.org/10.1101/2021.05.19.21257330
50. Khanna, A. *et al.* (2015) Microstates in resting-state EEG: current status and future directions. *Neurosci. Biobehav. Rev.* 49, 105–113
51. O'Neill, G.C. *et al.* (2018) Dynamics of large-scale electrophysiological networks: a technical review. *NeuroImage* 180, 559–576
52. Newson, J.J. and Thiagarajan, T.C. (2019) EEG frequency bands in psychiatric disorders: a review of resting state studies. *Front. Hum. Neurosci.* 12, 521
53. Baker, A.P. *et al.* (2014) Fast transient networks in spontaneous human brain activity. *eLife* 3, e01867
54. Tibon, R. *et al.* (2021) Transient neural network dynamics in cognitive ageing. *Neurobiol. Aging* 105, 217–228
55. Colclough, G.L. *et al.* (2016) How reliable are MEG resting-state connectivity metrics? *NeuroImage* 138, 284–293
56. Hasson, U. *et al.* (2010) Reliability of cortical activity during natural stimulation. *Trends Cogn. Sci.* 14, 40–48
57. Ben-Yakov, A. and Henson, R.N. (2018) The hippocampal film editor: sensitivity and specificity to event boundaries in continuous experience. *J. Neurosci.* 38, 10057–10068
58. Geerligs, L. and Campbell, K.L. (2018) Age-related differences in information processing during movie watching. *Neurobiol. Aging* 72, 106–120
59. Kauttonen, J. *et al.* (2018) Brain mechanisms underlying cue-based memorizing during free viewing of movie *Memento*. *NeuroImage* 172, 313–325
60. Meshulam, M. *et al.* (2021) Neural alignment predicts learning outcomes in students taking an introduction to computer science course. *Nat. Commun.* 12, 1922
61. Chen, J. *et al.* (2017) Shared memories reveal shared structure in neural activity across individuals. *Nat. Neurosci.* 20, 115–125
62. Lenormand, D. and Piolino, P. (2022) In search of a naturalistic neuroimaging approach: exploration of general feasibility through the case of VR-fMRI and application in the domain of episodic memory. *Neurosci. Biobehav. Rev.* 133, 104499
63. Zadbood, A. *et al.* (2017) How we transmit memories to other brains: constructing shared neural representations via communication. *Cereb. Cortex* 27, 4988–5000
64. Roberts, G. *et al.* (2019) Towards OPM-MEG in a virtual reality environment. *NeuroImage* 199, 408–417
65. Seymour, R.A. *et al.* (2022) Interference suppression techniques for OPM-based MEG: opportunities and challenges. *NeuroImage* 247, 118834
66. Liebherr, M. *et al.* (2021) EEG and behavioral correlates of attentional processing while walking and navigating naturalistic environments. *Sci. Rep.* 11, 22325
67. Krugliak, A. and Clarke, A. (2022) Towards real-world neuroscience using mobile EEG and augmented reality *Sci. Rep.* 12, 2291
68. Hasson, U. and Honey, C.J. (2012) Future trends in neuroimaging: neural processes as expressed within real-life contexts. *NeuroImage* 62, 1272–1278
69. Nastase, S.A. *et al.* (2020) Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage* 222, 117254
70. Maguire, E.A. (2012) Studying the freely-behaving brain with fMRI. *NeuroImage* 62, 1170–1176
71. Hamilton, L.S. and Huth, A.G. (2020) The revolution will not be controlled: natural stimuli in speech neuroscience. *Lang. Cogn. Neurosci.* 35, 573–582
72. Matusz, P.J. *et al.* (2019) Are we ready for real-world neuroscience? *J. Cogn. Neurosci.* 31, 327–338
73. Sonkusare, S. *et al.* (2019) Naturalistic stimuli in neuroscience: critically acclaimed. *Trends Cogn. Sci.* 23, 699–714
74. Haxby, J.V. *et al.* (2020) Naturalistic stimuli reveal a dominant role for agentic action in visual representation. *NeuroImage* 216, 116561
75. Finn, E.S. and Bandettini, P.A. (2021) Movie-watching outperforms rest for functional connectivity-based prediction of behavior. *NeuroImage* 235, 117963
76. Kurby, C.A. and Zacks, J.M. (2018) Preserved neural event segmentation in healthy older adults. *Psychol. Aging* 33, 232–245
77. Michelmann, S. *et al.* (2021) Moment-by-moment tracking of naturalistic learning and its underlying hippocampo-cortical interactions. *Nat. Commun.* 12, 5394
78. Chang, C.H.C. *et al.* (2021) Relating the past with the present: information integration and segregation during ongoing narrative processing. *J. Cogn. Neurosci.* 33, 1106–1128
79. Willems, R.M. *et al.* (2020) Narratives for neuroscience. *Trends Neurosci.* 43, 271–273
80. Greene, D.J. *et al.* (2018) Behavioral interventions for reducing head motion during MRI scans in children. *NeuroImage* 171, 234–245
81. Pashler, H. and Wagenmakers, E. (2012) Editors' introduction to the special section on replicability in psychological science: a crisis of confidence? *Perspect. Psychol. Sci.* 7, 528–530
82. Simmons, J.P. *et al.* (2011) False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* 22, 1359–1366
83. Ioannidis, J.P.A. (2005) Why most published research findings are false. *PLoS Med.* 2, e124
84. Niso, G. *et al.* (2021) Good scientific practice in MEEG research: progress and perspectives. *OSF Preprints* Published online September 29, 2021. https://doi.org/10.31219/osf.io/n2ryp
85. Turner, B.O. *et al.* (2018) Small sample sizes reduce the replicability of task-based fMRI studies. *Commun. Biol.* 1, 62
86. Poldrack, R.A. *et al.* (2017) Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nat. Rev. Neurosci.* 18, 115–126
87. Szucs, D. (2016) A tutorial on hunting statistical significance by chasing *N*. *Front. Psychol.* 7, 1444
88. Reagh, Z.M. *et al.* (2020) Aging alters neural activity at event boundaries in the hippocampus and posterior medial network. *Nat. Commun.* 11, 3980
89. Cooper, R.A. *et al.* (2021) Mapping the organization and dynamics of the posterior medial network during movie watching. *NeuroImage* 236, 118075
90. Campbell, K.L. *et al.* (2016) Robust resilience of the frontotemporal syntax system to aging. *J. Neurosci.* 36, 5214–5227
91. Zacks, J.M. *et al.* (2001) Human brain activity time-locked to perceptual event boundaries. *Nat. Neurosci.* 4, 651–655
92. Ben-Yakov, A. and Dudai, Y. (2011) Constructing realistic engrams: poststimulus activity of hippocampus and dorsal striatum predicts subsequent episodic memory. *J. Neurosci.* 31, 9032–9042