Neuron
NeuroView

Interactionist Neuroscience

David Badre,[1,2,4,*] Michael J. Frank,[1,2,4] and Christopher I. Moore[1,3,4]
[1]Brown Institute for Brain Science, Brown University, Providence, RI 02912-1978, USA
[2]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI 02912-1978, USA
[3]Department of Neuroscience, Brown University, Providence, RI 02912-1978, USA
[4]Authors are listed in alphabetical order
*Correspondence: david_badre@brown.edu
http://dx.doi.org/10.1016/j.neuron.2015.10.021

We argue that bidirectional interaction between animal and human studies is essential for understanding the human brain. The revolution in meso-scale study of circuits in non-human species provides a historical opportunity. However, to fully realize its potential requires integration with human neuroscience. We describe three strategies for successful interactionist neuroscience.

Introduction

In the early part of the 20th century, physics was in the middle of a revolution. Novel methods for measuring infinitesimal particles and macro-scale forces were generating unique new data. Computational advances, such as the theory of relativity, provided new frameworks for understanding these findings. Put simply, physics was the place to be for scientific revolutions in 1915.

One hundred years later, neuroscience is having a similar moment. The revolution in genetic engineering and determination of the molecular and cellular machinery underlying neural communication has led to several recent Nobel prizes. Advances in molecular and cellular biology have paved the way for understanding fundamentals of the nervous system, from neurogenesis to programmed cell death to circuit dynamics, based on the study of invertebrates such as flies, worms, and crabs. Such models are essential for continued innovation in understanding biology and in testing methods for broader application across phylogenetic levels.

At the other end of the spectrum, the invention of methods for imaging human brains non-invasively has provided a transformative advance in discovering basic features of its systems-level organization, a macro-scale road map for understanding perception, action, and cognition. The recent over billion-euro "Human Brain Project" and subsequent United States "BRAIN" Initiative have called for a similar revolution at the meso level between these two scales, focusing investigation on the activity and connectivity of extensive systems of neurons. This effort is well served by a variety of recent innovations in genetic targeting of cell types and new optical methods that allow control and recording from large numbers of specific neurons in animal models, particularly mice.

Scientific interest in understanding this meso scale—how circuits and systems of areas compute to serve biological information processing—has many motivations. First, such operations are interesting in their own right, as an example of how biological elements can work together to process information. Such information provides not only insight into biology, but also inspiration for artificial intelligence and computational architectures.

Second, the meso scale is particularly informative for basic understanding of how complex human cognition emerges from neural computation. Viewed from a clinical perspective, the goal of improving human brain health requires understanding not only the micro scale of what genes and epigenetic factors underlie normal function, but how these factors alter the network-level interactions that are widely regarded as the proximal neural events underlying behavior. On the basic science side, if we seek to fundamentally understand the human condition—why we are who we are—then the meso level of analysis is an essential part of the answer that science can provide, and we have a unique new opportunity to pursue it.

Yet, investigation at the meso scale cannot provide the whole answer. If one is concerned with human brain function—whether as a funding agency, as a research institution, or as a scientist—concrete links must be made between animal model systems and the complex human system. Our view is that these links can only be achieved by active, direct interaction between human and non-human neuroscience research. Failing to substantiate that meso-scale neural processes in animal models parallel those underlying human functions can easily lead to false inferences. Conversely, failing to ground human neuroscience in basic biological principles—only addressable in animal models—may result in descriptive abstractions that lead to theoretical dead-ends. Yet, these direct links seldom occur spontaneously and must be directly sought. We elaborate these issues and propose specific steps to increase the probability of such connections.

Why Isolated Work at the Meso Level Is Insufficient if the Goal Is Understanding Human Brain Function in Health and Disease

As stated above, the meso-scale revolution provides an exciting opportunity; it is hard to overstate the potential gains that will come from in vivo work in animal models at all levels. However, a common assumption is that animal models provide an example, albeit in miniature, of the same principles that apply for analogous computations in humans. This linear or "strong" reductionist assumption is often flawed for many reasons.

First, emergence of more complex behavior from simpler systems—even if we understood all the rules in such

systems—is often deeply challenging. As a colloquial example, consider the game of chess. An individual can readily learn all the rules of chess, as there are only a few dozen. Yet, the emergent tree of games arising from these rules is so vast that human players have never exhausted its branches. Indeed, it is only by having a relatively detailed model of the typical course of a game—and, in the case of experts, the strategies they employ—that one can understand and make predictions about the behavior that will unfold in any specific instance (Holland, 1999).

Such attempts to infer more complex systems from reductionistic information present an even thornier challenge in biology, where we are unsure of the rules themselves, much less their complex, emergent realization. A salient example comes from genetics. We have sequenced the entire human genome, yet the complex biochemical interactions that the system undergoes to produce specific features in a developed organism are still challenging to predict. The rapidly growing field of epigenetics is largely predicated on addressing this problem. Thus, understanding such complex systems is crucially dependent on identifying the properties of the existing system (watching a game of chess or fully characterizing a phenotypic expression) and then trying to describe how rules combine to produce that system.

The emergence problem holds within strictly human neuroscience as well. Even if we could feasibly conduct meso-level research in the human brain, one would still have to understand how lower-level biological variables (e.g., spike rates in neuronal subtypes) interact with each other, how they are influenced by higher-level variables (e.g., coordinated synchrony across regions), and within a dynamically changing environment to explain everyday human functioning. Animal neuroscience is often posed as a means of studying the base-level biological principles underlying human brain function. So, it becomes important in these cases to demonstrate their explanatory power directly in the complex, human system. Doing so requires coordination with human neuroscience.

As Norbert Wiener and Arturo Rosenblueth famously said, ''the best material model for a cat is another, preferably the same one.'' While the logical challenges to emergence reflected in this quote and in our above discussion are not cause for defeatism in the use of reduced models, they emphasize the need for interaction. The need for interaction is supported not only by logic, but also by extensive data, as discussed in the next section.

Wide-Ranging Examples of How Inference from Mouse-Meso to Humans Can Fail: Clinical Trials

In addition to the preceding logical challenge, there is substantial quantitative data directly supporting the prediction that biological principles determined only in animal models often do not predict human responses, even when those models are probed in detail at the meso level. Phase II clinical trials are designed to test treatments in humans that show pre-clinical efficacy in animal models. Often, these treatments derive from new insights into the mechanistic biology of disease processes in mice. Moreover, their efficacy in animal models and promise is such that they lead corporations and funding agencies to spend the millions of dollars required to proceed to human testing. In short, they have to be highly promising based on these initial screens to make it to this level of analysis.

Despite the prior thought and financial investment required in such trials, they fail far more often than not. Between 2008 and 2010, the success rate of such trials was only 18%. Approximately 50% of this failure rate was due to a lack of clinical efficacy in humans, despite the well-researched success in animal models (Arrowsmith, 2011). Put simply, vetting biology extensively, and introducing large financial risk in treatment development, still does not produce effective predictions of how a drug will fare in humans.

The challenge of translational work is of course not a reason to abandon animal research. To gain the most value from the indispensable contributions of animal models, we must have a firm understanding of their translation to the human. This translation requires active, coordinated human neuroscience research. We note that similar arguments have been raised by scientists in defense of the need for human stem cell lines in addition to animal lines.

Why Only Focusing on Human Neuroscience Fails Logically and Misses a Historic Scientific Moment

Just as focusing exclusively at the meso scale is fundamentally limiting, transformative progress cannot be made simply by studying the emergent, complex system in humans. With an MRI scanner at almost every institution, it is easy to forget that the advent of accessible, non-invasive human neuroimaging and neurostimulation techniques was a historic revolution in neuroscience. Yet, currently available techniques have well-known and significant limitations, not the least of which is that they largely stop at the macro scale. As such, these methods provided limited data for a circuit-level of description, and they are unable to address questions related to cell-type-specific functions within a region. These differences can be critical to understanding function, predicting behavior and its modulations by pharmacological agents. Thus, many of the biological premises on which these complex systems are based require testing in animal models at multiple levels.

Why Cooperate Now?

A commonly asked question is whether investments in progress in human neuroscience should wait until further breakthroughs are made at the meso level. This question denies logic in several ways. Chief among these is that without an integrated approach, we cannot know if progress at the meso level applies in the human, for the reasons stated above. Similarly, there is much to gain from understanding the principles that govern human cognition and behavior captured by quantitative models, even if nothing is known about neural mechanisms in advance. For example, behavioral economic models have provided constraints on the utility functions employed in human decision making and the ways in which they handle uncertainty and delays (e.g., hyberbolic discounting). This approach allows the animal researcher to identify relevant targets to be explained in terms of underlying mechanism. Indeed, there are already several examples for which cognitive and computational theory derived from human cognitive and neural science has positively influenced animal research (some other prominent

examples are described in Frank and Badre, 2015).

A further practical consequence of self-segregation into human- or animal-only research programs is that without conscious attempts to align efforts and progress, human and non-human work will continually diverge into their own idiosyncratic subdomains and will increasingly lose common ground for communication, making prospects for their eventual synthesis increasingly remote. Perversely, the "parallel" approach to human and non-human research has, in some cases, placed them indirectly in competition for resources and focus within the scientific community, and so progress in one is seen as a threat to continued progress in the other.

The Path to Full Cooperative Neuroscience: Computation, Cross-Species Parallelism, and Cross-Cultural Immersion

It follows that directly coordinated human-animal neuroscience should be a key paradigm going forward. However, for practical, philosophical, and sociological reasons, human and non-human work typically do not spontaneously merge. Rather, achieving the necessary level of interaction requires direct focus by the scientific community to overcome significant gaps in knowledge, resources, and prejudicial attitudes. We highlight three key areas where the necessary links can be made: (1) computation, (2) cross-species parallelism, and (3) cross-cultural immersion.

Theory Is the Key: Next-Generation Computational Models Blending Algorithmic Insights and Biophysically Precise Circuits

Computational theoretical neuroscience typically occurs at one of many levels. Models seek to understand the "nuts and bolts" details of neurophysiology in single cells, the dynamics among multiple neurons and across circuits in brain systems, or the computations needed for perception, action, and cognition at an abstract level. A formal separation between these levels is often assumed to be the appropriate strategy when devising theory, given the multiple ways computational and algorithmic-level processes can be realized and the fact that neurons and circuits, and their emergent processes, are interesting in their own right.

A fruitful new strategy is modeling that formally combines these levels and/or attempts to link across them (for detailed discussion and review, see Frank, 2015). These models are structured to test algorithmic understanding, but seek to use real details of the biology, to test whether the computations proposed at the abstract level can be performed by more realistic circuits, loops, and hierarchies. There are many advantages to this approach. First, because such models respect the algorithmic level, the predictions they make at the meso level are ideal for guiding experimentation. Even if a mouse is not capable of all the features of complex human cognition that initially motivated the model, the neural details of the model that were necessary to realize the algorithmic objective can be tested in more precise ways. For example, models of cortico-striatal loops underlying working memory and hierarchical cognitive control functions in primates (e.g., Chatham and Badre, 2015) can be used to guide and/or interpret dissection of their implementational details in the mouse sensori-motor loop—even if mice are not using this neural architecture to apply complex abstract rules or hierarchical executive control.

Second, details of the biology can provide clues to the algorithmic level. For example, algorithmic models of reinforcement learning (RL) inherited from computer science and psychology suggest that a key signal needed to drive learning is a reward prediction error (the difference between expected and obtained reward value). These models motivated the seminal discovery that phasic changes in midbrain dopamine firing convey reward prediction errors (Montague et al., 1996). This finding overturned the previous theory that dopaminergic signals were either purely motor or purely reward related (as opposed to an error signal needed for learning) and has received converging support across methods and species.

Nevertheless, the details of the neural implementation of reward prediction error have also reciprocally informed the algorithmic level. A closer look at the biology of the dorsal striatum reveals that this system comprises two neural populations that respond in opponent ways to dopamine (due to their differential expression of D1 and D2 receptors) and that have differential effects on action selection (due to differential projections to distinct basal ganglia output nuclei). Neural models that incorporate these details predict a wide variety of empirical effects of dopamine manipulations on reward learning and choice that are not accommodated by classical RL models without such opponent processes (Collins and Frank, 2014). These model predictions have been corroborated by pharmacological, genetic, patient, and imaging studies in humans, and more precise evidence for the models' proposed roles for separate D1 and D2 populations have been confirmed using optogenetic and other genetic engineering methods (for review, see Collins and Frank, 2014). These dynamic circuit models of corticostriatal interactions have further facilitated a connection between algorithmic models of learning (typically concerned with dynamics across trials) and those of decision making (concerned with dynamics of choice processes within trials). This example illustrates how theoretical neuroscience can provide a framework for bidirectional interaction, providing explanatory power at both levels.

Similar synergistic approaches across levels have been successfully applied to other aspects of decision making (attractors models and sequential sampling models from mathematical psychology; Wang, 2012), episodic memory (pattern separation and completion in the hippocampus; O'Reilly and McClelland 1994), and working memory (prefrontal cortical gating networks and biophysical models of active maintenance; Cohen et al., 2002). All of these cases of bidirectional interaction have led to advances in empirical work across methods and species while also allowing biophysical and functional models to inform and refine each other.

A crucial second target for computation to link animal and human studies are models focused on translating patterns of biologically specific activity into the signals measured in humans. As one example, recent studies have provided precise and formulated explanations for magnetoencephalography (MEG) signals based on details of the biophysics and

connectivity of neocortical columns (Jones et al., 2007). Such "translators" provide a rational and mechanistic test of whether the macro-neurophysiological signatures observed in humans can be generated by the detailed neural processes observed in animal models. Formal explanations for the neural drivers of EEG and fMRI signals are further away, though strong recent progress has been made, and translators that solve for these signals are similarly essential. This requirement for the success of intersectional neuroscience agrees with recent calls for a research focus on determining these solutions (Devor et al., 2013).

Building the Meso-Macro Mesh: Explicit Alignment of Mouse, Monkey, and Human Studies

Research in humans and animal models must be intentionally aligned, so that they can be mutually informative. First, research must be directed to identify strong functional homologies that are supported by more than superficial behavioral analogy. Since the cognitive revolution in the late 20th century, we have known that merely observing a similar behavior between two species (or even two individuals of the same species) does not mean that the underlying processing is the same. Nevertheless, behavioral analogy is often the only bar that is met when linking human and animal models, risking time and resources on potentially irrelevant and idiosyncratic lines of research. Thus, efforts should be made to draw strong functional homologies between levels. As described above, such homologies can occur at an algorithmic level, whereby human and animal models are aligned within a particular theoretical framework. The previously described example of corticostriatal loop models of working memory provides such a case.

The use of parallel methods that allow collection of identical types of data is another means of drawing functional homologies. Neuroimaging methods are a key focus, as many techniques driving the meso-mouse revolution are not available in the human. Notable among these is fMRI, a key engine of human cognitive neuroscience and a method actively being pursued in monkeys and mice. By conducting animal studies that parallel human studies, we can directly test whether the same areas or networks are activated, and the same activity patterns within them. If observed, such parallelism goes a long way to substantiating the inference that monkey electrophysiological findings will apply in the human. Similar fMRI studies can be conducted in mouse, where the entire toolbox of leading edge techniques, particularly optogenetic control, can be applied.

Advances in the analysis of complex EEG and surface-potential signals similarly offer an opportunity to draw functional homologies in functional spectral signatures between animals and humans. The recent rapid progress in detailed and systematic quantitative analysis of human single-neuron and local field potential recordings also provides a potentially crucial translational tool. While these data are only obtained in subjects with advanced clinical symptomology, they can provide a key bridge between levels, in addition to providing an important area in advancement of potential treatment modalities.

Moving forward, research tools for intersectional work require continuing innovations that allow us to test formal predictions and to provide a substrate for functional homologies between systems. For example, gains in temporal and spatial resolution of human and non-human neuroscience methods (like primate or mouse fMRI), as well as better understanding of convergent methods, will permit focus on those functional signatures that can guide non-human work. Similarly, new statistical tools for handling large datasets and formally testing complex, biologically plausible computational models would help leverage gains made in formal theory and would increase the range of testable hypotheses. Such parallelism and inferences made from it require deeper understanding of neuro-vascular communication. This active field of biological research will require the same kinds of computational translational "solving tools" to allow micro-scale inferences from the macro-scale fMRI signals (Devor et al., 2013).

Just as advances in animal imaging can help draw strong functional analogy between levels, human cognitive neuroscientists need to test frameworks that make connection to biological principles from animal studies. Human-level cognitive neuroscience has the ability to identify macro-level neural systems that contribute to human cognitive or behavioral function. Humans can perform tasks that no other animal can perform, and experiments involving humans leave considerably less doubt about their relevance to the human condition. Despite their well-known limitations, the fact that human neuroimaging methods, like fMRI, EEG, and MEG, provide these types of observations makes them indispensable to the progress of any neuroscience of the human brain. A prominent neurophysiologist once quipped, "All fMRI does is tell me where to put my electrodes." This comment was meant to be diminishing, but it actually highlights the great value in even the crudest degree of systems-level knowledge about the human brain. Of course, at its best, modern human cognitive neuroscience contributes much more than mere spatial localization (Poldrack and Farah, 2015).

Thus, an emphasis should be placed on careful, task-based human cognitive neuroscience research that is mechanistically informed and derives, where possible, from biologically grounded theoretical frameworks or deeply analogous animal research. The field of cognitive neuroscience already has many successful examples of this type of research. As a few examples, investigation of pattern completion/separation processes in the human hippocampus has driven the discovery of parallels in animal models (Yassa and Stark, 2011). fMRI studies of short-term memory in humans have complemented observations in animals indicating that individual items can be maintained based on dynamic forms of coding, as opposed to sustained delay period activity (D'Esposito and Postle, 2015). The previously discussed working memory gating mechanisms have derived inspiration from understanding of corticostriatal motor systems and neurotransmitter systems studied in animals, but have tested crucial extensions of these mechanisms to more complex forms of cognitive control in human beings (Chatham and Badre, 2015). Finally, the development of techniques to test predictions from computational models with fMRI, so-called model-based fMRI, holds great promise to both inform and be informed by the hybrid models described above.

In sum, these research programs in human cognitive neuroscience are already ongoing in certain sectors; we are encouraging their continued growth and widespread emphasis.

Training the Next Generation: Fully Integrated Immersion across Levels

We must train a new generation of neuroscientists that are question driven and not technique limited. This generation should be prepared to take full advantage of data from multiple sources and levels of analysis to address their research hypotheses.

The first and most important step in achieving this goal is to conceive of training as centered around questions rather than techniques or levels of analysis. Students should learn to propose the best hypotheses they can, and then find and use the methods that will allow them to prosecute these ideas. This view is in contrast to the notion that immersion in a level of analysis is the only goal of graduate education and that question-driven inquiry and learning can be a secondary priority that simply occurs en passant.

There are several implications that accompany this shift in perspective. Perhaps the most important implication is that training serially in multiple laboratories should be an option fully available to graduate and postdoctoral students. Students should join labs for projects, not necessarily for their educational life, and be able to conduct projects serially across different groups in the course of a thesis or postdoctoral training period as appropriate. The notion that an advanced student is best trained by only experiencing a single lab is provincial; sampling multiple styles and levels of analysis is a great way to not only achieve an intersectionist perspective, but also understand what approaches and ideas will best suit that particular student in their future career.

The most important shift to accommodate this goal is explicit funding mechanisms that divorce students from the indentured financial dependence of exclusively single-lab and single-level-of-inquiry mentorship, not just for a brief period of ''sampling'' such as rotations, but for an entire education. The one-lab/one-level-of-analysis mentality permeates graduate programs and is an almost obligatory paradigm of postdoctoral training. While every student is unique, promoting integration for a significant number of students is important to the future of neuroscience. Federal grants should exist that will allow students the lateral freedom to move between laboratories or facilitate collaborations across them. Dedicating a portion of NRSA and K-Award NIH grants to multi-lab mentorship, and explicitly encouraging integrated and multi-level training, would be a major step in this direction.

This training is essential for generating scientists who conduct their research program at multiple levels themselves, such as doing human and non-human research or modeling at the biological circuit and algorithmic levels. These individuals provide valuable links between communities, but their training is necessarily longer than that experienced by the scientists working exclusively at the human or non-human level.

Many graduate programs have elements that encourage such interactions, including rotations aimed at providing multi-level experience (not just auditions for potential full-time participation in a lab) and ''boot camps'' where several kinds of experimentation are taught simultaneously for an intensive period to all students. These efforts are laudable, and we hope they are expanded on, in addition to the more overt support for detailed and extensive training at multiple levels described above.

In addition to the spontaneous interactions across levels that accompany a boot camp experience, programs should be designed to encourage sustained contact among scientists at different levels of analysis. For a number of social and practical reasons, such contact rarely happens spontaneously. For example, the degree to which human-level cognitive neuroscientists and those pursuing meso- to molecular-level neuroscience have contact is highly variable across training programs, particularly as scientists in these disciplines are trained in diverse graduate programs, ranging from psychology to medicine to biology. This limited understanding makes it difficult for these scientists to incorporate complex concepts gained from these disciplines into their own research programs. One means of enhancing contact would be to house graduate student and postdoctoral offices outside the lab and organize students by theme or question rather than laboratory.

In addition to the obvious need for coursework at each level of analysis, contact and sophistication across levels can also be driven by individual courses focused on specific questions and spanning levels, incorporating human, animal, and computational principles. Examples of topics that can be taught this way might include reinforcement learning theory, neural dynamics, decision making, or memory models. Such teaching will likely require multiple instructors from different levels of expertise, and the act of coordinating such courses is in itself a driver toward interactionist principles, as such collaboration in teaching can often yield meaningful intellectual exchanges and collaborations between faculty.

A macro-level requirement needed to meet each of these specific sub-goals is a learning environment that is not only borderless (allowing students to explore beyond one lab), but a truly integrated training program that spans from low-level molecular models to high-level cognitive science. Achieving this goal might require a re-envisioning of graduate education, away from the cantonization that typifies many programs.

Conclusion

Recent advances in meso-level neuroscience are both exciting and timely. The next few years will see exciting discoveries emerge from a new focus on this level of inquiry. However, when left on their own, the molecular, systems (meso), or human levels of neuroscience will not spontaneously converge to provide understanding of complex human functions as diverse as language, planning, emotion, decision making, and memory. Innovation and progress must evolve in interaction between these levels, bridged by formal modeling and identification of deep functional homologies. Further, training the next generation of neuroscientists must encourage openness and sophistication at multiple levels of analysis. The strategy of waiting for connections between levels to spontaneously emerge is not a viable one; if we seek the deepest and richest

understanding of human brain function, we need to actively commit to synergy.

ACKNOWLEDGMENTS

We would like to thank Karla Kaun, Gilad Barnea, and Anne Hart for helpful feedback on an earlier version of this viewpoint.

REFERENCES

Arrowsmith, J. (2011). Nat. Rev. Drug Discov. *10*, 328–329.

Chatham, C.H., and Badre, D. (2015). Current Opinion in Behavioral Sciences *1*, 23–31.

Cohen, J.D., Braver, T.S., and Brown, J.W. (2002). Curr. Opin. Neurobiol. *12*, 223–229.

Collins, A.G.E., and Frank, M.J. (2014). Psychol. Rev. *121*, 337–366.

D'Esposito, M., and Postle, B.R. (2015). Annu. Rev. Psychol. *66*, 115–142.

Devor, A., Bandettini, P.A., Boas, D.A., Bower, J.M., Buxton, R.B., Cohen, L.B., Dale, A.M., Einevoll, G.T., Fox, P.T., Franceschini, M.A., et al. (2013). Neuron *80*, 270–274.

Frank, M.J. (2015). Linking across levels of computation in model-based cognitive neuroscience. In An Introduction to Model-Based Cognitive Neuroscience, B.U. Forstmann and E. Wagenmakers, eds. (New York: Springer).

Frank, M.J., and Badre, D. (2015). Cognition *135*, 14–20.

Holland, J.H. (1999). Emergence: From Chaos to Order (New York: Basic Books).

Jones, S.R., Pritchett, D.L., Stufflebeam, S.M., Hämäläinen, M., and Moore, C.I. (2007). J. Neurosci. *27*, 10751–10764.

Montague, P.R., Dayan, P., and Sejnowski, T.J. (1996). J. Neurosci. *16*, 1936–1947.

O'Reilly, R.C., and McClelland, J.L. (1994). Hippocampus *4*, 661–682.

Poldrack, R.A., and Farah, M.J. (2015). Nature *526*, 371–379.

Wang, X.J. (2012). Curr. Opin. Neurobiol. *22*, 1039–1046.

Yassa, M.A., and Stark, C.E.L. (2011). Trends Neurosci. *34*, 515–525.