Disentangling the skeins of brain

Brian Knutson[1,2], Tara Srirangarajan[1]

1. Department of Psychology, Stanford University, Stanford, CA USA

2. Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA USA

Final version (22.11.21):

Accepted for publication in the Journal of Cognitive Neuroscience

Keywords: emergence, reduction, expansion, causality, incentive

Contact: knutson'at'stanford.edu

Abstract

Some have argued that the brain is so complex that it cannot be understood using current reductive approaches. Drawing on examples from decision neuroscience, we instead contend that combining new neuroscientific techniques with reductive approaches that consider central brain components in time and space has generated significant progress over the past two decades. This progress has allowed researchers to advance from the scientific goals of description and explanation to prediction and control. Resulting knowledge promises to improve human health and well-being. As an alternative to the extremes of reductive versus emergent approaches, however, we propose a middle way of “expansion.” This expansionist approach promises to leverage the specific spatial localization, temporal precision, and directed connectivity of central neural components to ultimately link levels of analysis.

“Time is what keeps everything from happening at once

(and space is what keeps everything from happening to me).”

– Ray Cummings (elaborated by Arthur Dudden)

Students often learn that science has four primary goals, which include not only observation and explanation, but also prediction and control (Watson, 1913). By successively meeting these goals, an investigator can infer and establish causal influence. Further, the degree to which a theoretical account meets these goals can serve as a measure of its scientific merit. Below, we compare how well a new emergent account versus an older reductive account meet these scientific goals. We conclude, however, by suggesting that another hybrid “expansionist” account may best address all four scientific goals, and ultimately link levels of analysis.

Thesis: Emergence

Summarizing the thematic thrust of his book, *The Entangled Brain* (Pessoa, 2022), Luiz Pessoa argues that scientists should strive to understand the brain as an emergent complex system (Pessoa, 2023). To do so, he advocates focusing on the “interactional complexity” of the brain with respect to three criteria: 1) pervasive anatomical connectivity, 2) distributed functional coordination, and 3) networks as units of analysis. Extrapolating from this emergent neuroscientific approach, he further proposes dissolving boundaries between apparently distinct psychological functions (e.g., perception, emotion, cognition, motivation, and so forth).

By focusing on emergence rather than reduction, Pessoa proffers an intriguing alternative to more conventional analytic approaches by defining brain function as arising from interactions of components which are not reducible to the activity of those components (or “nondecomposable”). Emergence and reduction could be visualized as occupying opposite ends of a spectrum. At one extreme, emergence implies bidirectional connections between all components within a level of analysis (e.g., brain or behavior), but not necessarily across levels of analysis. Thus, some interaction of components at one level might influence the interaction of components at an adjacent level of analysis, but how this occurs is left unspecified. At the other extreme, reduction implies that all components at a higher (e.g., behavior) level are connected to (or equivalent to) components at an adjacent lower level (e.g., brain), and so can be reduced to that level of analysis (Figure 1).

Antithesis: Reduction

But is the call to abandon reduction premature? Since the turn of the twenty-first century and in combination with technical innovations, reductive analyses have supported substantial advances in understanding brain function. For example, new hybrid fields have emerged (e.g., affective neuroscience and decision neuroscience), which combine neuroimaging with largely reductive analyses to demonstrate that incentives can drive brain activity, which then reliably predicts choice in individuals as well as groups (e.g., Knutson & Greer, 2008). To reconsider the potential benefits versus costs of reductive approaches, we next revive discarded analogies and invert

proposed criteria (Pessoa, 2022). To provide concrete illustrations while resisting the seduction of pure philosophical speculation, we anchor these reconsiderations to empirical findings culled from our area of expertise – the neuroscience of incentive processing.

Vivid analogies of the brain as tangled skeins of yarn or as a car with interconnected functioning components are invoked but then discarded as inaccurate (Pessoa, 2023). We believe, however, that these simplifying analogies still help to illuminate brain function. The analogy of tangled skeins echoes through the writing of early neurophysiologists (e.g., the “enchanted loom” of Sherrington), but continues to aptly describe any random slice of brain visualized at sufficiently precise resolution. Neuroscientists still seek to “disentangle” these neural threads using increasingly precise methods which have advanced from early staining tools (yielding the neuron doctrine and Nobel Prizes; Cajal, 1954) to recent tissue clearing and viral tracing techniques (e.g., Kim et al., 2017). The analogy of a vehicle with functionally distinct but connected components also remains helpful for predicting not only the effects of local lesions, but also disconnections between critical components, and even the focus of interventions. For instance, clinicians are currently using diffusion tractography to more accurately place deep-brain stimulators designed to treat debilitating conditions ranging from movement disorders to addiction (Krauss et al., 2021). By disentangling the skeins of brain, researchers may advance neuroscientific interventions that can help repair the vehicle of the mind.

Beyond discarding useful analogies, abandoning reductive analysis also seems premature. If emergence and reduction define opposite ends of a spectrum, emergent principles could be inverted to instead generate reductive principles, which could then be evaluated. We consider each of three inverted principles, along with applications to incentive processing, below:

Centers (versus connectivity)

A reductive principle of centers might retain the notion that critical components are more central (or even necessary) to support a given function than other components, while still acknowledging their interconnection. For instance, the architecture of frontostriatal circuits implies that components are connected in an "ascending spiral" that facilitates translation of motivation into motion (e.g., Haber & Knutson, 2010). Not all of these components support the same functions, however, and not all of their connections are bidirectional. Extensive evidence now indicates that component function depends on locale, with ventromedial components subserving more motivational functions and dorsolateral components subserving more motoric and control functions (Voorn et al., 2004). Additionally, while the frontal cortex sends massive direct glutamatergic projections to the striatum, projections returning from the striatum to the frontal cortex are indirect and more bidirectional, first coursing through the pallidum and thalamus. This distinct topologically directed architecture helps to channel and focus the flow of neural impulses through interconnected corticostriatal components. Ignoring the architecture of critical functional centers can have consequences for interpreting neural data. For instance, recently adopted multiband

Functional Magnetic Resonance Imaging (FMRI) acquisition sequences can induce high-frequency noise, especially in the ventromedial components of these corticostriatal loops, which may obscure visualization of motivational function but not motor function (Srirangarajan et al., 2021). Omitting the function of these central components thus runs the risk of biasing findings, not only in single studies, but also throughout the broader neuroimaging literature.

Time (versus coordination)

A reductive principle of time might posit that central components are recruited at specific points in time to support relevant functions. By extension, even in an interconnected circuit, all components are not active at the same point in time or in response to the same demands. For example, during the initial development of an incentive processing task for FMRI (i.e., the Monetary Incentive Delay or MID task), task trials were modeled as a whole (collapsing across both anticipation and outcome phases), generating results which suggested that monetary incentives increased activity in the dorsal striatum and dorsomedial prefrontal cortex (Knutson et al., 2000). Subsequent separate modeling of anticipation and outcome phases of each trial, however, indicated that while anticipation of reward increased activity in the ventral striatum (including the NAcc), receipt of reward outcomes instead increased activity in the Medial PreFrontal Cortex (MPFC) (Knutson et al., 2001). Multiple meta-analyses now confirm the replicability of this temporally distinct recruitment profile of different mesolimbic regions in response to reward anticipation versus outcomes (e.g., Knutson & Greer, 2008; Liu et al., 2011; Oldham et al., 2018). This temporal distinction also

extends to specifying computational models. For instance, models of updates in reward prediction (also called “reward prediction errors”) in response to both reward cues and outcomes correlated with activity in both the NAcc and the MPFC. Separate models of only reward cue updates correlated primarily with NAcc activity, however, while models of only reward outcome updates instead correlated primarily with MPFC activity (Knutson & Wimmer, 2007). By extension, the longer timescale (often ranging from 10-20 minutes) of most resting state and even task-based correlational analyses (sometimes called “functional connectivity”) produces results indicating that NAcc and MPFC activity are robustly correlated (Chen et al., 2022), but can obscure lags or conditional specificity of those correlations. Based on data modeled over a longer timescale, one might mistakenly infer that functionally dissociable components of the mesolimbic circuit constitute a single functional network (e.g., the “Default Mode Network”). Ignoring temporal precision thus runs the risk of obscuring recruitment of distinct connected regions at different points in time.

Space (versus networks)

A reductive principle of space might posit that critical components are spatially delimited, as is their connection with other components. Several of the brain structures implicated in incentive processing are relatively small, irregularly shaped, and lie below the cortex. For example, the human NAcc occupies approximately 980 cubic mm on either side of the brain (Neto et al., 2008), has an elongated tubular shape which extends forward from the middle of the brain, and lies near the base of the brain. In neuroimaging analyses, while spatial smoothing can reduce the significance required to

detect an effect, excessive spatial smoothing (e.g., > 8 mm FWHM) systematically shifts NAcc activation foci towards the back of the brain and even into adjacent structures (e.g., the putamen; Sacchet & Knutson, 2012). Worse yet, extensive clustering criteria intended to reduce the significance required to detect an effect may simply exclude activity in regions smaller than the requisite cluster size (e.g., > ~900 cubic mm for the NAcc). Disregarding spatial structure thus runs the risk of compromising the detection of central functional components.

To summarize, inverting the proposed principles of emergent analysis supports a reconsideration of the value of reductive analyses. In recent years, reductive analyses have conferred several benefits on decision neuroscience. First, the principle of centers has allowed researchers to localize neural activity that scales with the expected value of stimuli and then to use that activity to predict choice. Second, the principle of time has supported researchers in honing temporal resolution to deconstruct different phases of experimental tasks, revealing that distinct centers play different functional roles in response to incentive cues versus outcomes. Third, the principle of space has helped researchers to focus on regions of different sizes and shapes to demonstrate the interplay of those connected centers in processing incentives. Further, reductive analyses have minimized potential costs related to functional confounds, mismatched spatial resolution, and incommensurate temporal resolution (either during acquisition or analysis). Thus, the benefit to cost ratio of reductive analysis in neuroimaging continues to add value. The benefit to cost ratio of emergent analysis, however, remains to be determined. Fortunately, researchers can do better than “everything, everywhere, all at

once.” Considering interconnected centers situated in space and time has empowered researchers to disentangle the skeins of brain.

Synthesis: Expansion

While the inverted principles above imply support for reductive over emergent analyses, we believe that forcing a choice between the two presents a false dichotomy. Instead, we have suggested a middle way that combines elements of both, dubbed “expansion” (Knutson & Srirangarajan, 2019). Consistent with a “deep science” framework for linking levels of analysis, the expansionist approach initially seeks to identify and link central components at adjacent levels of analysis (where “link” implies directional and causal influence at a similar timescale). Once a robust link across levels has been established, strong connections to central components within levels can be identified. For example, in humans, FMRI activity in the NAcc at the brain level can predict choice seconds later at the behavioral level (e.g., Kuhnen & Knutson, 2005). These levels of analysis may also contain sublevels that can be linked. Specifically, the brain level might include chemical and anatomical sublevels, while the behavior level might include self-reported experiential and choice sublevels (Figure 2). For example, in rats, optogenetic stimulation of midbrain dopamine neurons increases FMRI activity in the NAcc, which predicts individuals’ willingness to work to self-administer that stimulation (Ferenczi et al., 2016).

Expansive analyses can add value to reductive and emergent analyses by allowing researchers not only to address the first two goals of science (description and

explanation) but also the second two (prediction and control), and thereby to establish a causal influence. The need for causal accounts is not only theoretical – it is also practical. For example, by detecting and interfering with an electrophysiological correlate of reward anticipation in the NAcc (i.e., low frequency potentials in the delta band), mice and humans can be diverted from compulsive consumption of high fat foods, and possibly in the future, substances of abuse (Shivacharan et al., 2022; Wu et al., 2018).

In summary, the goals of science do not end with observation and explanation, but also extend to prediction and control. Scientists should not settle for less. We do not agree that “causal explanations […] miss the point” (Pessoa, 2023). Combining new techniques and expansionist analyses can lead us to the brink of confirming causality. Now is the time to forge ahead, not to turn away.

Figures

Figure 1. Schematics of multilevel components and connections illustrating reduction, expansion, and emergence accounts.

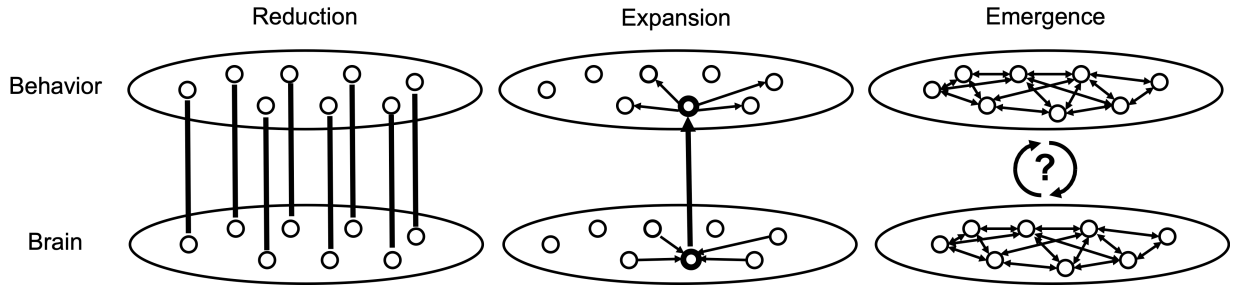
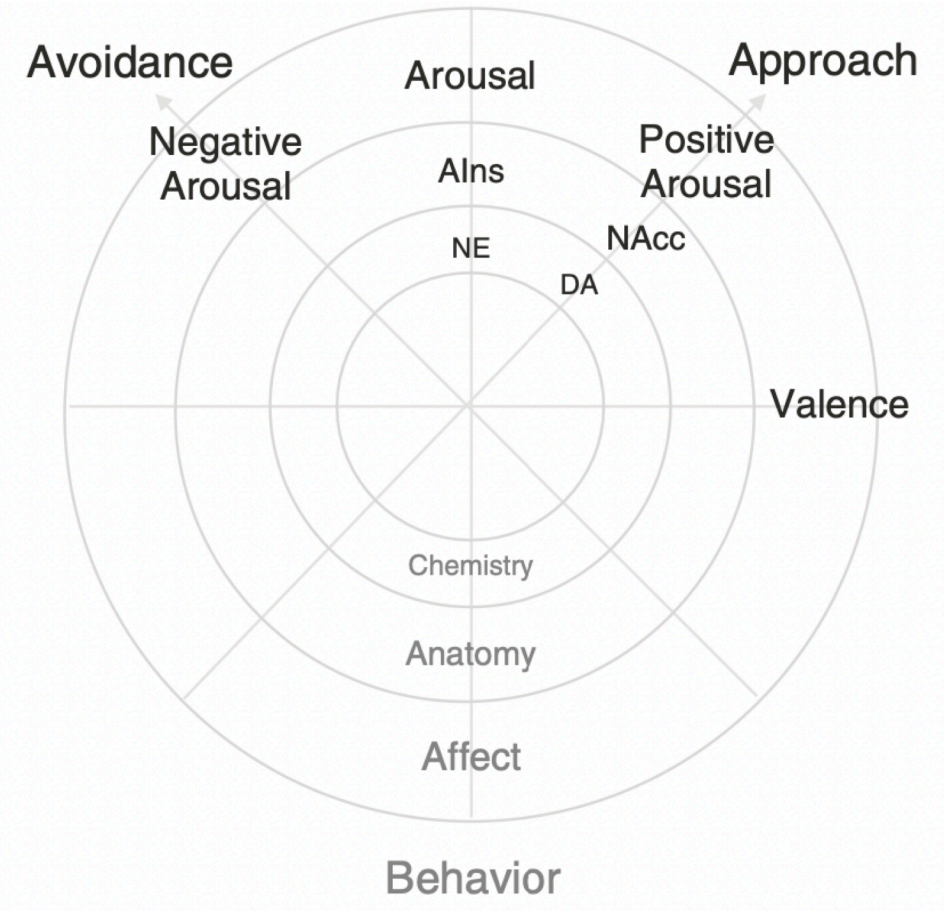


Figure 2. Proposed causal mapping of components across multiple levels of analysis (adapted from Knutson et al., 2014). Causal influence flows from central to outer circles at a similar timescale. Abbreviations: DA = DopAmine, NE = NorEpinephrine, NAcc = Nucleus Accumbens, AIns = Anterior Insula.



References

Cajal, R. y. (1954). *Neuron theory or reticular theory? Objective evidence of the anatomical unity of nerve cells.* Consejo Superior de Investigaciones Cientificas.

Chen, J., Tam, A., Kebets, V., Orban, C., Ooi, L. Q. R., Asplund, C. L., Marek, S., Dosenbach, N. U. F., Eickhoff, S. B., Bzdok, D., Holmes, A. J., & Yeo, B. T. T. (2022). Shared and unique brain network features predict cognitive, personality, and mental health scores in the ABCD study. *Nature Communications*, *13*(1), 2217. https://doi.org/10.1038/s41467-022-29766-8

Ferenczi, E. A., Zalocusky, K. A., Liston, C., Grosenick, L., Warden, M. R., Amatya, D., Katovich, K., Mehta, H., Patenaude, B., Ramakrishnan, C., Kalanithi, P., Etkin, A., Knutson, B., Glover, G. H., & Deisseroth, K. (2016). Prefrontal cortical regulation of brainwide circuit dynamics and reward-related behavior. *Science*, *351*(6268). https://doi.org/10.1126/science.aac9698

Haber, S. N., & Knutson, B. (2010). The reward circuit: Linking primate anatomy and human imaging. *Neuropsychopharmacology*, *35*(1), 4–26. https://doi.org/10.1038/npp.2009.129

Kim, C. K., Adhikari, A., & Deisseroth, K. (2017). Integration of optogenetics with complementary methodologies in systems neuroscience. *Nature Reviews Neuroscience*, *18*(4), Article 4. https://doi.org/10.1038/nrn.2017.15

Knutson, B., Fong, G. W., Adams, C. M., Varner, J. L., & Hommer, D. (2001). Dissociation of reward anticipation and outcome with event-related fMRI. *NeuroReport*, *12*(17), 3683–3687. https://doi.org/10.1097/00001756-200112040-00016

Knutson, B., & Greer, S. M. (2008). Anticipatory affect: Neural correlates and consequences for choice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1511), 3771–3786. https://doi.org/10.1098/rstb.2008.0155

Knutson, B., Katovich, K., & Suri, G. (2014). Inferring affect from fMRI data. *Trends in Cognitive Sciences*, *18*(8), 422–428. https://doi.org/10.1016/j.tics.2014.04.006

Knutson, B., & Srirangarajan, T. (2019). Toward a deep science of affect and motivation. In *Nebraska Symposium on Motivation* (Vol. 66, pp. 193–220). Springer. https://doi.org/10.1007/978-3-030-27473-3_7

Knutson, B., Westdorp, A., Kaiser, E., & Hommer, D. (2000). FMRI visualization of brain activity during a monetary incentive delay task. *NeuroImage*, *12*(1), 20–27. https://doi.org/10.1006/nimg.2000.0593

Knutson, B., & Wimmer, G. E. (2007). Splitting the Difference. *Annals of the New York Academy of Sciences*, *1104*(1), 54–69. https://doi.org/10.1196/annals.1390.020

Krauss, J. K., Lipsman, N., Aziz, T., Boutet, A., Brown, P., Chang, J. W., Davidson, B., Grill, W. M., Hariz, M. I., Horn, A., Schulder, M., Mammis, A., Tass, P. A., Volkmann, J., & Lozano, A. M. (2021). Technology of deep brain stimulation: Current status and future directions. *Nature Reviews Neurology*, *17*(2), Article 2. https://doi.org/10.1038/s41582-020-00426-z

Kuhnen, C. M., & Knutson, B. (2005). The neural basis of financial risk taking. *Neuron*, *47*(5), 763–770. https://doi.org/10.1016/j.neuron.2005.08.008

Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional

neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *35*(5), 1219–1236. https://doi.org/10.1016/j.neubiorev.2010.12.012

Neto, L. L., Oliveira, E., Correia, F., & Ferreira, A. G. (2008). The human nucleus accumbens: Where is it? A stereotactic, anatomical and magnetic resonance imaging study. *Neuromodulation: Journal of the International Neuromodulation Society*, *11*(1), 13–22. https://doi.org/10.1111/j.1525-1403.2007.00138.x

Oldham, S., Murawski, C., Fornito, A., Youssef, G., Yücel, M., & Lorenzetti, V. (2018). The anticipation and outcome phases of reward and loss processing: A neuroimaging meta-analysis of the monetary incentive delay task. *Human Brain Mapping*, *39*(8), 3398–3418. https://doi.org/10.1002/hbm.24184

Pessoa, L. (2022). *The Entangled Brain*. The MIT Press.

Pessoa, L. (2023). <title> JoCN, [this issue].

Sacchet, M. D., & Knutson, B. (2012). Spatial smoothing systematically biases the localization of reward-related brain activity. *NeuroImage*, 66, 270–277. https://doi.org/10.1016/j.neuroimage.2012.10.056

Shivacharan, R. S., Rolle, C. E., Barbosa, D. A. N., Cunningham, T. N., Feng, A., Johnson, N. D., Safer, D. L., Bohon, C., Keller, C., Buch, V. P., Parker, J. J., Azagury, D. E., Tass, P. A., Bhati, M. T., Malenka, R. C., Lock, J. D., & Halpern, C. H. (2022). Pilot study of responsive nucleus accumbens deep brain stimulation for loss-of-control eating. *Nature Medicine*, *28*(9), Article 9. https://doi.org/10.1038/s41591-022-01941-w

Srirangarajan, T., Mortazavi, L., Bortolini, T., Moll, J., & Knutson, B. (2021). Multi-band FMRI compromises detection of mesolimbic reward responses. *NeuroImage*, *244*, 118617. https://doi.org/10.1016/j.neuroimage.2021.118617

Voorn, P., Vanderschuren, L. J. M. J., Groenewegen, H. J., Robbins, T. W., & Pennartz, C. M. A. (2004). Putting a spin on the dorsal-ventral divide of the striatum. *Trends in Neurosciences*, *27*(8), 468–474. https://doi.org/10.1016/j.tins.2004.06.006

Watson, J. B. (1913). Psychology as the behaviourist views it. *Psychological Review*, *20*(2), 158–177. https://doi.org/10.1037/h0074428

Wu, H., Miller, K. J., Blumenfeld, Z., Williams, N. R., Ravikumar, V. K., Lee, K. E., Kakusa, B., Sacchet, M. D., Wintermark, M., Christoffel, D. J., Rutt, B. K., Bronte-Stewart, H., Knutson, B., Malenka, R. C., & Halpern, C. H. (2018). Closing the loop on impulsivity via nucleus accumbens delta-band activity in mice and man. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(1), 192–197. https://doi.org/10.1073/pnas.1712214114