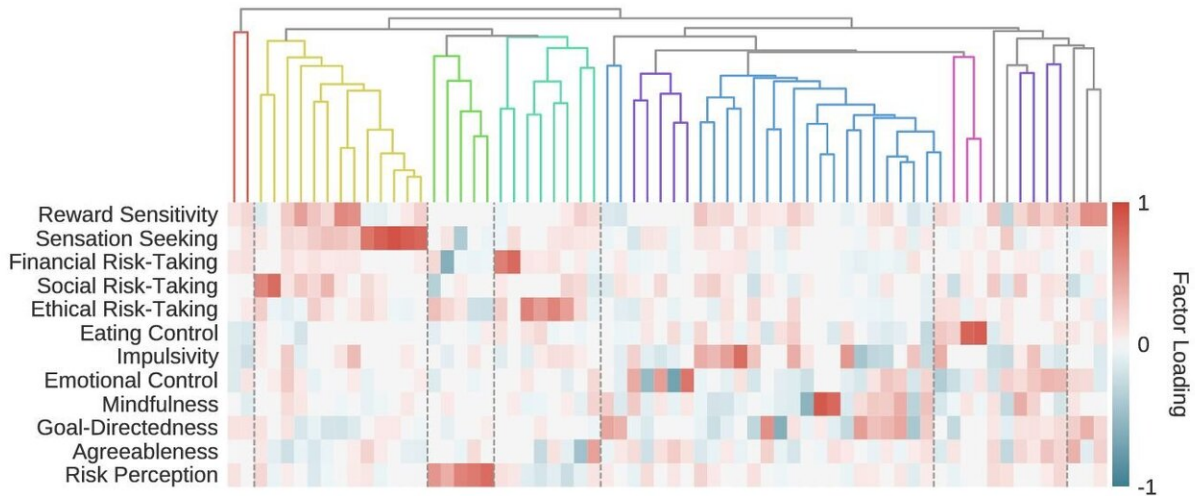


Raising the standard for psychology research

August 6 2019



Survey ontology. 66 survey dependent variables were projected onto 12 factors discovered using exploratory factor analysis, represented by the heatmap. Rows are factors and columns are separate dependent variables ordered based on the dendrogram above. Credit: TACC

In recent years, efforts to understand the workings of the mind have taken on new-found urgency. Not only are psychological and neurological disorders—from Alzheimer's disease and strokes to autism and anxiety—becoming more widespread, new tools and methods have emerged that allow scientists to explore the structure of, and activity within, the brain with greater granularity.

The White House launched the BRAIN Initiative on April 2, 2013, with the goal of supporting the development and application of innovative technologies that can create a dynamic understanding of brain function. The initiative has supported more than \$1 billion in research and has led to new insights, new drugs, and new technologies to help individuals with brain disorders.

But this wealth of research comes with challenges, according to Russell Poldrack, a psychology professor with a computing bent at Stanford University. Psychology and neuroscience struggle to build on the knowledge of its disparate researchers.

"Science is meant to be cumulative, but both methodological and conceptual problems have impeded cumulative progress in psychological science," Poldrack and collaborators from Stanford, Dartmouth College and Arizona State University wrote in a *Nature Communications* paper out in May 2019.

Data Archivist

Part of the problem is practical. With hundreds of research groups undertaking original research, a central repository is needed to host and share data, compare and combine studies, and encourage data reuse. To address this curatorial challenge, in 2010 Poldrack launched a platform called OpenfMRI for sharing fMRI studies.

"I'd thought for a long time that [data sharing](#) was important for a number of reasons," explained Poldrack, "for transparency and reproducibility and also to help us aggregate across lots of small studies to improve our power to answer questions."

OpenfMRI grew to nearly a hundred datasets, and in 2016 was subsumed into OpenNeuro, a more general platform for hosting brain

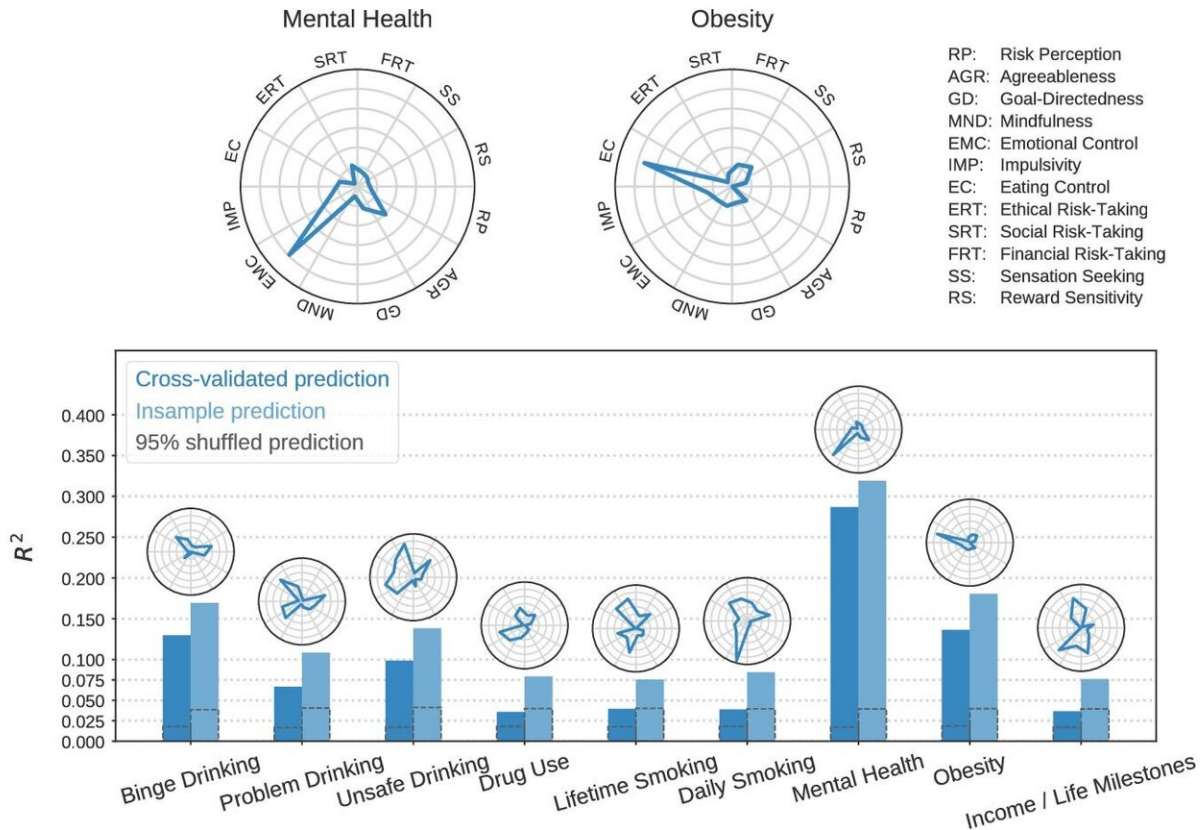
imaging studies. That platform today has more than 220 datasets, including some like "The Stockholm Sleepy Brain Study" and "Neural Processing of Emotional Musical and Nonmusical Stimuli in Depression," that have been downloaded hundreds of times.

Brain imaging datasets are relatively large and require a large repository to house them. When he was developing OpenfMRI, Poldrack turned to the Texas Advanced Computing Center (TACC) at The University of Texas at Austin to host and serve up the data.

A grant from the Arnold Foundation allowed him to host OpenNeuro on Amazon Web Services for a few years, but recently Poldrack turned again to TACC and to other systems that are part of the NSF-funded Extreme Science and Engineering Discovery Environment (XSEDE) to serve as the cyberinfrastructure for the database.

Part of the success of the project is due to the development of a common standard, BIDS—Brain Imaging Data Structure (BIDS)—that allows researchers to compare and combine studies in an apples-to-apples way. Introduced by Poldrack and others in 2016, it earned near-immediate acceptance and has grown into the lingua franca for neuroimaging data.

As part of the standard creation, Poldrack and his collaborators built a web-based validator to make it easy to determine whether one's data meets the standard.



Prediction of target outcomes using survey factor scores, estimated from 2500 shuffles of the target outcome. Ontological fingerprints displayed as polar plots indicate the standardized beta value for each significant survey factor. The ontological fingerprint for the two best predicted outcomes are reproduced at the top. Credit: Ian W. Eisenberg, Patrick G. Bissett, A. Zeynep Enkavi, Jamie Li, David P. MacKinnon, Lisa A. Marsch & Russell A. Poldrack

"Researchers convert their data into BIDS format, upload their data and it gets validated on upload," Poldrack said. "Once it passes the validator and gets uploaded, with a click of a button it can be shared."

Data sharing alone is not the end goal of these efforts. Ultimately, Poldrack would like to develop pipelines for computation that can

rapidly analyze brain imaging datasets in a variety of way. He is working with the CBrain project, based at McGill University in Montreal, Canada, to create containerized workflows that researchers can use to perform these analyses without requiring a lot of advanced computing expertise, and independent of what system they are using.

He is also working with another project called BrainLife.io based at Indiana University, which uses XSEDE resources, including those at TACC, to process data, including data from OpenNeuro.

Many of the datasets from OpenNeuro are now available on BrainLife, and there is a button on those datasets that takes one directly to the relevant page at BrainLife, where they can be processed and analyzed using a variety of scientist-developed apps.

"In addition to sharing data, one of the things that having this common data standard affords us is the ability to automatically analyze data and do the kind of pre-processing and quality control that we often do on imaging data," he explained. "You just point the container at the data set, and it just runs it."

Rethink Discipline-Wide Assumptions

Things would be simple if formatting, storage, and sharing were the only problems the field faced. But what if the common methods researchers used for analyzing studies introduced biases and errors, leading to a lack of reproducibility? Moreover, what if the underlying assumptions about the way the mind worked were fundamentally flawed?

A study published in 2018 in *Nature Human Behaviour* that sought to replicate 21 social and behavioral science papers from *Nature* and *Science* found that only 13 could be successfully replicated. Another meta-study under the auspices of the Center for Open Science, re-ran 28

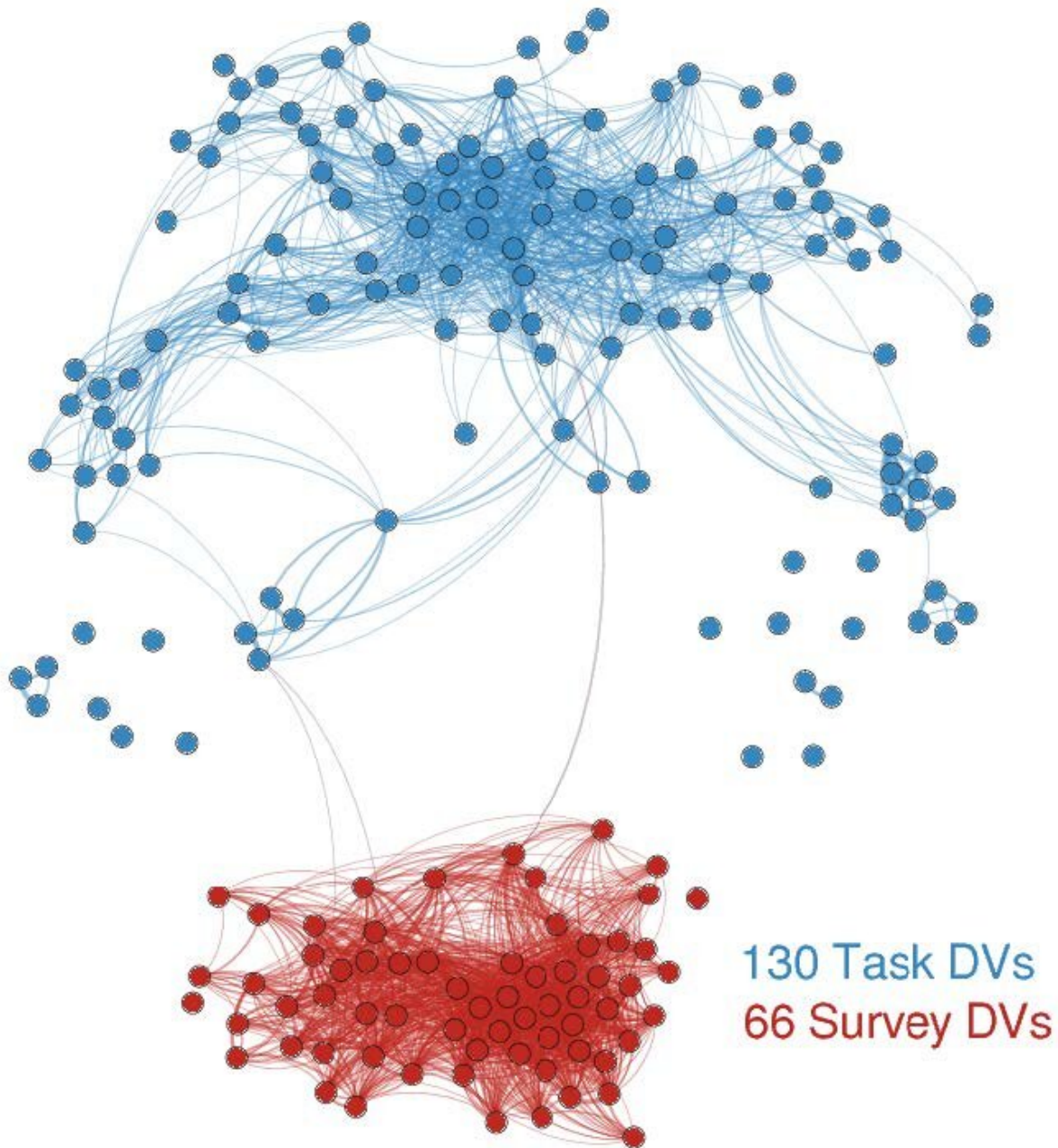
classic and contemporary studies in psychology and found that 14 failed to replicate. This has led to retroactive suspicions about decades worth of results.

Poldrack and his collaborators addressed both the methodological and assumption problems in their recent *Nature Communications* paper by applying more rigorous statistical methods to try to uncover the underlying structures of the mind, a process they call 'data-driven [ontology](#) discovery.'

Applying the approach to studies of self-regulation, the researchers tested the ability of survey questionnaires and task-based studies to predict an individual's likelihood of being at risk for alcoholism, obesity, drug abuse, or other self-regulation-related issues.

In their study, 522 participants took 23 self-report studies and performed 37 behavioral tasks. From each of these 60 measures, the team derived multiple dependent variables thought to capture psychological constructs. Using the dependent variables, the team first tried to create "a psychological space"—a way of quantifying the distance between dependent variables to determine how various types of behavior that are often seen as separate cluster or correlate to each other. They used these "ontological fingerprints" to determine the contribution of various psychological constructs to the final predictive model.

The statistical approach used in the study, and enabled by supercomputers at TACC, goes far beyond the standard methods used in typical psychological studies.



Psychological graph of all dependent variables (DVs). Graphical lasso was used to estimate a sparse undirected graph representing the relationships amongst all DVs. Nodes represent DVs while edges represent partial correlation between two DVs (thickness reflects strength). Credit: Ian W. Eisenberg, Patrick G. Bissett, A. Zeynep Enkavi, Jamie Li, David P. MacKinnon, Lisa A. Marsch & Russell A. Poldrack

"We're bringing to bear serious machine learning methods to determine what's correlated with what, and what has generalizable predictive accuracy, using methods that are still fairly new to this area of research," Poldrack said.

They found that some predicted targets, like mental health and obesity, had simple ontological fingerprints, such as "emotional control" and "problematic eating," but that other fingerprints were more complicated. They also found that task-based studies—common in psychological research—had almost no predictive ability.

"I'm always leery of saying our research will be useful for diagnosis, but it almost certainly will be useful for a better understanding of how to do diagnosis and the underlying functions that relate to certain outcomes, like smoking or problem drinking or obesity," Poldrack said.

Motivating the effort is a re-examination of the way that we talk about mental illness.

"Breaking these disorders into diagnostic categories like schizophrenia, bipolar disorder, or depression, is just not biologically realistic," he said. "Both genetics and neuroscience show that those disorders have way more overlap in terms of their genetics and their neurobiology, than differences. So, I think that there are new paradigms that might emerge that would be helped by a better understanding of the brain."

High performance computing allows researchers to apply much more sophisticated methods to determine knowledge distributions and figure out how significant results are.

"We can use sampling techniques to randomize the data 5,000 times and

re-run big models many times," Poldrack said. "That's not realistically possible without supercomputers."

It used to be the case that the progress of science was dependent on the ability to create a molecule or synthesize a chemical. But increasingly progress in science depends on the ability to ask the right question about a big data set, and then to be able to actually feasibly get an answer to that question.

"And," said Poldrack, "there's a lot of questions that, without high performance computing, you can't feasibly get an answer to."

Despite the crises of faith that has struck the field in recent years, Poldrack believes [psychological science](#) has a lot to say that is very reliable about why humans do what they do, and that neuroscience gives us ways to understand where that comes from.

"We're trying to understand really complex things," he said. "It has to be realized that everything we say is probably wrong, but the hope is that it can get us a little bit closer to what's right."

More information: Ian W. Eisenberg et al, Uncovering the structure of self-regulation through data-driven ontology discovery, *Nature Communications* (2019). [DOI: 10.1038/s41467-019-10301-1](https://doi.org/10.1038/s41467-019-10301-1)

Provided by University of Texas at Austin

Citation: Raising the standard for psychology research (2019, August 6) retrieved 20 April 2024 from <https://medicalxpress.com/news/2019-08-standard-psychology.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private

study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.