

Psychological Science Needs A Seat at the Informatics Table

February 29, 2012

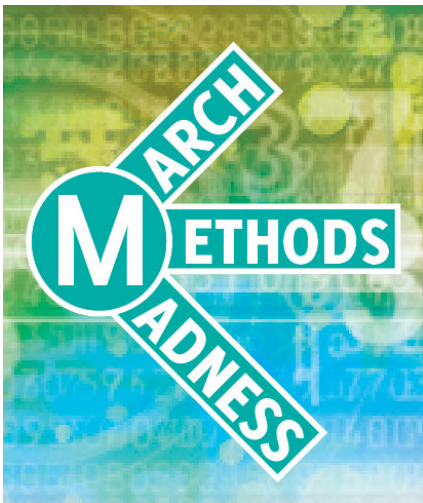


Tal Yarkoni

Scientific progress depends on our ability to harness and apply modern information technology. Many advances in the biological and social sciences now emerge directly from advances in the large-scale acquisition, management, and synthesis of scientific data. The application of information technology to science isn't just a happy accident; it's also a field in its own right — one commonly referred to as *informatics*. Prefix that term with a Greek root or two and you get other terms like *bioinformatics*, *neuroinformatics*, and *ecoinformatics* — all well-established fields responsible for many of the most exciting recent discoveries in their parent disciplines.

Curiously, following the same convention also gives us a field called *psychoinformatics* — which, if you believe Google, doesn't exist at all (a search for the term returns only 500 hits as of this writing; Figure 1). The discrepancy is surprising, because labels aside, it's clear that psychological scientists are already harnessing information technology in powerful and creative ways — often reshaping the very way we collect, organize, and synthesize our data. Take the very notion of a psychological science “lab.” It used to imply a physical space in which researchers collect data from human participants; yet there are now many psychological scientists who rarely conduct conventional lab-based studies and instead rely on Web-based samples. And why not? As several studies have shown, online samples are typically more diverse, larger, and cheaper to acquire than university samples. While there are many things one can't do online (e.g., shock people or scan their brains), the availability of an effectively limitless userbase has fundamentally altered the data collection landscape in many areas of psychological science.

Or consider the way our increasingly networked lives offer unprecedented opportunities for psychological investigation. In the simple act of living, many of us generate a continuous stream of information: We text our friends, track our locations with GPS, upload pictures of our activities, and stream music and movies through the air. Thanks to modern technology, much of this information persists indefinitely in our Gmail, Facebook, and Netflix accounts. And psychological scientists are discovering that the mundane residue of our everyday lives makes extraordinary data. Recent studies have harnessed data from online services to study the structure of social networks, the way personality shapes bloggers' writing, and the evolution of people's mood over the course of the day, to name just a few applications. Such efforts are undoubtedly just scratching the surface; an unprecedented wealth of data on people's activities, relationships, thoughts, and feelings is increasingly available to us for research purposes — if we just know where and how to look.



Then there's mobile technology. We're told there's an app for everything, and psychological science is no exception. In the past few years, we've seen psychological scientists use mobile phones to study everything from daydreaming to visual word recognition. Ironically, new mobile technologies are already rendering slightly less new technologies obsolete. A few years ago, Matthias Mehl and colleagues at the University of Arizona developed an innovative recording device called the EAR (Electronically Activated Recorder). People wear the EAR while they're out and about, and the device periodically records audio snippets of their daily lives. As a long string of publications attests, data from the EAR has taught us some very interesting things about what people do when they're not in the lab. Yet the EAR is arguably already obsolete — deprecated in favor of the iEAR, an iPhone app that does exactly the same thing, but runs on hardware many people already own. This trend underscores the growing role of passive data collection in psychological science: People are already generating masses of data voluntarily and effortlessly, so why not put it to good use?

Of course, no matter how we obtain our data, we still need to synthesize it into a meaningful form. And here again, psychological scientists can benefit from approaches pioneered in other informatics disciplines. For instance, much of bioinformatics has focused on developing databases and software for large-scale genomic data mining. Admittedly, psychological data aren't as structured as genetic data; we can't yet sequence mental states using just four DNA bases. Nonetheless, there are plenty of opportunities for more modest applications. For example, my colleagues and I recently developed a platform called Neurosynth (<http://neurosynth.org>; Yarkoni et al., 2011) that uses data reported in thousands of published neuroimaging articles to automatically generate fMRI meta-analyses and “decode” mental states from patterns of brain activity. From a technical standpoint, our work offers little

in the way of innovation; the text mining and machine learning techniques we used have been around for decades. But by applying old techniques to a new domain, we were able to develop tools that offer neuroimaging researchers new insights into brain and cognitive function. A similar approach could potentially help address problems in many other areas of psychological science.

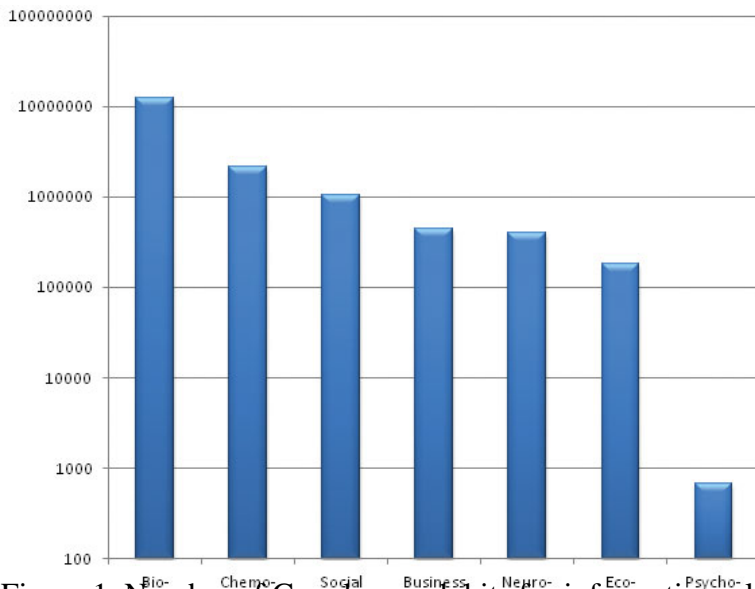


Figure 1. Number of Google search hits for informatics-related terms, by prefix.

Lastly, an increased emphasis on technological applications could help improve procedural aspects of psychological science research. As a number of commentators have recently pointed out in APS journals, psychological science, like any science, is an imperfect enterprise. Replication is under-emphasized, reporting standards can be lax, and peer review can be unreliable. There's a tremendous opportunity to develop tools and platforms that can help address such problems. One promising recent example is PsychFileDrawer (<http://psychfiledrawer.org>), an online archive of replication attempts in experimental psychology. Further down the road, one can envision a broad range of applications — for example, automated quality-control algorithms that detect patterns suggestive of publication bias (e.g., too many p values just below $p < .05$ in a researcher's published output), new publishing models that emphasize post- rather than pre-publication review, or text mining tools that automatically identify unexpected relationships between findings in different areas of psychology.

Importantly, our collective ability to develop platforms that take on the challenges and opportunities discussed above will require us to train researchers with skill sets that cut across traditional disciplinary boundaries. These researchers — call them psychoinformaticians — will have a strong grounding both in substantive psychological science and in relevant areas of the information and computer sciences. They'll build platforms to acquire and mine rich new datasets on an unprecedented scale. They'll develop tools for efficiently exploring and synthesizing huge amounts of information while minimizing the opportunity for human bias. And they'll come up with new ways of reporting and disseminating psychological findings to other scientists and to the public at large. If the trajectory of fields like biology and neuroscience is any guide, such developments will pay enormous dividends. We should embrace the emergence of psychoinformatics as a full-fledged discipline and work to ensure that psychological science remains a vibrant, forward-looking field, ready to benefit from technological advances as soon as they emerge.