Neuron
NeuroView

Taking Psychiatry Research Online

Claire M. Gillan[1,2,3,*] and Nathaniel D. Daw[4]
[1]Department of Psychology, New York University, 6 Washington Place, New York, NY, 10003, USA
[2]Department of Psychology
[3]Behavioural and Clinical Neuroscience Institute
University of Cambridge, Downing Site, Cambridge CB2 3EB, UK
[4]Department of Psychology and Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA
*Correspondence: Claire.gillan@gmail.com
http://dx.doi.org/10.1016/j.neuron.2016.06.002

Psychiatry is in need of a major overhaul. In order to improve the precision with which we can treat, classify, and research mental health problems, we need bigger datasets than ever before. Web-based data collection provides a novel solution.

Introduction

Unlike the rest of medicine, psychiatry has no objective diagnostic tests, instead relying entirely on self-report symptoms to classify and treat patients. While this approach has been useful in determining treatments for some people, most have an incomplete, and many an absent, response to treatment. We need to do better. An important part of progress toward that goal is establishing clear and robust links between clinical symptoms and the underlying neurobiological dysfunction (what physicians refer to as the etiology), so that one day, objective tests and targeted treatments can replace symptom-based differential diagnosis.

Yet research that might help to uncover these links is itself hampered by the same diagnostic difficulties. The Diagnostic and Statistical Manual of Mental Disorders (DSM), the standard psychiatric dictionary, draws categorical lines between disorders in a way that research has shown lacks biological specificity. First, these disorders are highly correlated, such that meeting the criteria for one disorder substantially increases the likelihood of being simultaneously diagnosed with another (termed "co-morbidity"). Second, these disorder categories are highly heterogeneous in terms of their defining symptomatology—meaning two patients with the same label often look very different. Typically, patients need meet only a subset of a long list of symptoms to get a diagnosis. In some cases, like attention deficit hyperactivity disorder (ADHD) and schizophrenia, two individuals can have the same diagnosis with absolutely no overlapping symptoms.

Given these taxonomic issues, when researchers use these categories as independent variables in research, it is unsurprising that the results lack specificity. For instance, in comparing cases to controls it is quite difficult to know if observed differences are attributable to a disorder of interest, versus one of the many comorbid disorders likely present in the case group. Perhaps for this reason, seemingly promising leads about etiology can grow hazy with further examination. Cognitive functions that were once linked to one disorder are evidently seen in many disorders, and the same is true of promising genetic associations and neural markers.

All these challenges point to a need for data-driven changes to our system of psychiatric classification and for accompanying changes in research methodology. In particular, vast datasets are almost certainly required to tease apart the contributions of comorbid and heterogeneous symptoms, to identify the aspects of clinical phenomenology that are the most biologically valid, to separate environmental from genetic contributions, to and handle high-dimensional problems such as predicting treatment response for drugs that have vastly different receptor affinity profiles.

Recognizing these limitations, scientists are increasingly turning to "big data" where possible. Kendler and colleagues have conducted extensive diagnostic interviews on thousands of patients in the search for more natural diagnostic boundaries that can explain patterns of comorbidity and shared familial risk across disorders (Kendler et al., 2011). Others have pooled existing datasets from different research sites to conduct "mega-analyses" on imaging data in an effort to better control for sources of known heterogeneity (e.g., age, gender, comorbidity) across studies (de Wit et al., 2014). Even more ambitious yet are large multi-center projects that collect diagnostic, genetic, cognitive, and neural data in the same individuals, such as the IMAGEN project, which followed 2,000 14-year-old adolescents over time with the aim of identifying predictors of psychiatric disorders (Schumann et al., 2010). These approaches are yielding fascinating new insights, but given how logistically complex and expensive they can be, they are not for everyone.

Fortunately, the Internet now offers a timely alternative for big psychiatry.

Taking Psychiatry Online

In the face of some similar challenges of power and methodology, the field of psychology has taken to online data collection in a big way. Smartphone applications have been developed that can collect data on gamified versions of popular cognitive tasks in thousands of people. Examples include recent efforts to assess age-related spatial working memory decline (McNab et al., 2015) and how reward receipt relates to momentary changes in happiness (Rutledge et al., 2014). In addition to single-session experiments, repeated testing over several weeks is possible in an online format and has been conducted as part of "citizen scientist" campaigns run in collaboration with popular media websites. Thanks to Amazon, who in 2005 launched a

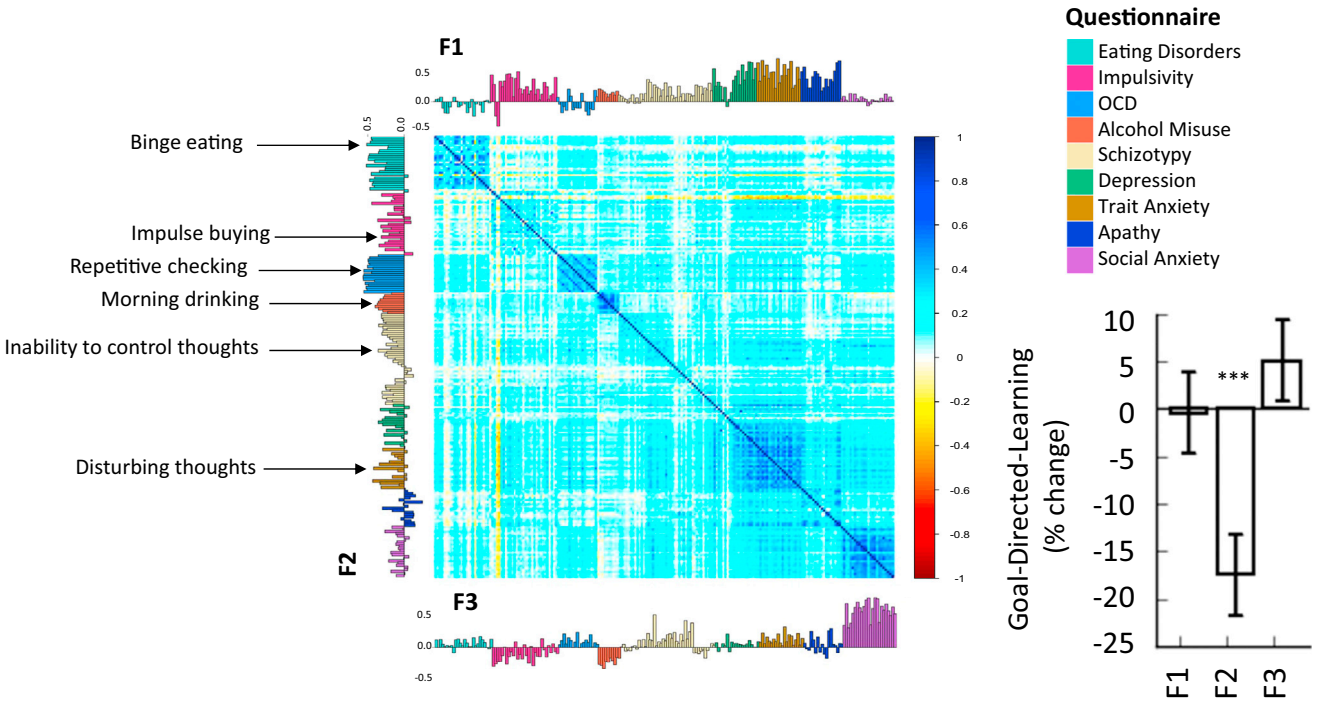

Figure 1. Big Data in Psychiatry Can Be Realized with Internet Testing

Figure reproduced from Gillan et al. (2016). Factor analysis of items from nine popular psychiatric self-report questionnaires in 1,413 individuals reveals a three-factor solution, comprising "Anxious-Depression" (F1), "Compulsive Behavior and Intrusive Thought" (F2), and "Social Withdrawal" (f3). Item descriptors from F2 (left) highlight how symptoms of compulsive behavior and intrusive thought can be found across questionnaires that purport to quantify discrete clinical phenomena. In an independent analysis, F2 was the only trans-diagnostic predictor of goal-directed deficits, which have been previously linked to the integrity of caudate function. The magnitude of this effect exceeded that of any of the nine original questionnaires that reflect more traditional psychiatric constructs. Error bars denote SEM.

popular crowdsourcing platform called Mechanical Turk, online data collection is now easier than ever, such that data can be simultaneously collected from hundreds of individuals per day.

This mode of data collection also has tremendous potential for psychiatric research, since the general population accessible via Turk spans the full range of psychiatric symptoms and severities (Shapiro et al., 2013). We tested the water recently, in a study that aimed to characterize a precise psychiatric symptom dimension based on its relation to an underlying neurocognitive mechanism (Gillan et al., 2016). Deficits in goal-directed control, which we have shown can cause people to get stuck in their habits (Gillan et al., 2015b), have been linked to compulsive behaviors in psychiatry such as addiction and OCD. While both theory and neurobiological data indicated that these deficits might be a specific source underlying compulsive aspects of psychiatry, recently published case-control studies have suggested otherwise—that these deficits are seen in many disorders, even those without compulsive features (Gillan et al., 2015a).

Suspecting this seeming lack of specificity might be down to the diagnostic categories rather than the brain-behavior links, we used web-based data collection to collect data from almost 2,000 subjects in order to carefully separate the contribution of different aspects of psychopathology to goal-directed control. Subjects in our study performed a behavioral task that assessed goal-directed control (one whose neurobiological foundations are also relatively well studied) (Daw et al., 2011) in their web browser. Instead of comparing one disorder to another, we examined self-report data also provided by subjects, regarding the symptoms and severity of nine different aspects of psychopathology espoused by DSM (depression, addiction, social anxiety, etc.). We first highlighted the problem—we showed that when we examine the data in a traditional way, assuming that, for example, eating disorder symptoms are independent of OCD symptoms, then the relationship between goal-directed control and psychiatric symptoms is indeed not specific to one disorder over another. This finding was of course predicted by the high rates of co-occurrence of disorders within individuals and the overlap of symptoms across disorders, described above.

Because we collected such a large sample of data, however, we were able to examine how specific symptoms (rather than disorders) naturally co-occur and test whether a dimensional approach to psychiatry might track the underlying cognitive construct better than the status quo. Specifically, we used factor analysis to look at the inter-correlation of subjects' answers to over 200 questions and found evidence for three trans-diagnostic psychiatric dimensions (Figure 1). Importantly, we found that these symptom dimensions possessed greater specificity in their relationship to our separate assessment of goal-directed control, compared to approaches that consider

eating disorders as distinct from OCD, for example. In line with predictions from basic science, a factor pertaining to compulsive behavior was a stronger predictor of goal-directed control deficits that any of the measures that quantify severity of DSM constructs, including OCD. This highlights how refining psychiatric classification using appropriately powered online samples can help us to make stronger links between neurocognitive mechanisms and symptoms—a critical step toward understanding the etiology of, and how best to treat, mental health problems.

Precision Therapeutics

Although the general population samples recruited via Turk are useful for answering many questions in psychiatry—for example, how symptoms might relate to neural or cognitive constructs that can be assessed with behavioral testing—other questions involving treatments and diagnosed patients are not so easily addressed this way. Nevertheless, other online testing approaches offer many opportunities.

Predicting which individuals will benefit most from a particular treatment is probably the most important, immediately accessible target for big, data-driven psychiatry. Indeed, computational approaches that require these large datasets have already shown promise in lab-based studies. For example, one study showed that electroencephalography (EEG) markers outperform clinical-determined treatment plans for depression (DeBattista et al., 2011). Other work has shown that a combination of baseline self-report measurements (e.g., comorbidity, age of onset) can be used to predict chronicity of depression (Kessler et al., 2016), and this might also inform drug treatment allocation (Chekroud et al., 2016). While carefully monitored, lab-based treatment studies are crucial, these are of course incredibly costly, time consuming, and as such have difficulty scaling up to collect the kind of sample sizes needed for carrying out predictive analyses. Online data collection can offer not only an appealing complement but also allow us to access difficult to reach populations that may never progress beyond primary care. For example, individuals who search the web for facts about antidepressants (a common practice before starting a new course) could be targeted via advertising on online forums as well as reputable patient information sites.

Hybrid approaches that rely on web-based testing, but recruit participants through clinical collaborators, may be even more effective. For example, advertising online studies through primary care centers, which are where most antidepressants are actually prescribed, would provide unprecedented access to vast numbers of treatment naive patients right before they start a new treatment.

Psychiatric Time Courses

In addition to improving patient access and sample sizes, online testing can facilitate much closer temporal monitoring of symptoms than is feasible for traditional laboratory studies. Smartphone apps can be used to record daily fluctuations in symptoms, which could be used to understand the time course of treatment response/non-response more precisely than even before. Cognitive behavioral therapy (CBT) providers are already working in this area—not only monitoring symptom fluctuations but also delivering novel and cost-effective CBT alternatives that show preliminary efficacy for substance use and depression (Donker et al., 2013). Such time course monitoring can also allow researchers to examine state versus trait dependence of cognitive and even neural markers of psychopathology. For example, apps that prompt users for monthly reports regarding their mood could be used to recall subjects for lab-based testing in periods of low versus high symptomatology to give new insights about cause and effect. Cognitive tests could of course be similarly administered without the need for patients to travel.

Biomarkers to Quantify Future Risk

Perhaps even more important than monitoring symptom fluctuations in diagnosed patients is identifying biomarkers that can predict who is at risk of developing a disorder in the future. This would permit earlier intervention, which some believe might improve long-term therapeutic outcome. The problem with this kind of research is that it requires the largest datasets of all. Scientists need to follow enormous numbers of healthy young individuals over many years—not only because just a small proportion will ultimately develop a mental health disorder, but also because within that subset we can expect multiple independent factors (genetic and environmental) to contribute to individual risk. The Internet is uniquely positioned to help recruit and maintain these large samples, but of course one limitation is that one can't acquire all of the data one would like via the web. While we can't collect brain scans or draw blood on-line, one way we can approximate the former is by using cognitive tests for which the neural correlates have been defined in pre-existing work (Gillan et al., 2016). In the absence of cognitive tasks that map onto well-defined neural circuits or genes, large-scale online testing can be used to identify reliable cognitive targets that can then be brought to smaller in-person samples where brain-imaging techniques might add further predictive power.

Field Experiments

Another way that the Internet changes the landscape of what is possible in psychiatric research is through the opportunity it provides us to conduct naturalistic experiments through popular social networking and search websites. For example, real-world experimental research that taps into learning, memory, and perception could be conducted on social media (leveraging website structure, clicks, etc.) and linked substance use, which can be inferred from Facebook "likes" (Kosinski et al., 2013). Other natural experiments have shown that gambling behavior (i.e., lottery sales) can be influenced by unexpected positive real-world events (Otto et al., 2016). As the popularity of online gambling grows, this provides a new opportunity for researchers to track fluctuations in real-world clinically relevant behaviors and identify risk factors for relapse. This approach is not without serious ethical implications, which came into the spotlight a couple of years ago when Facebook published a study in which they manipulated more than 650,000 users' "News Feeds" to omit positive or negative posts from other users, raising questions about informed consent in a commercial setting (Kramer et al., 2014). However, with clearly defined

consent parameters, these issues can be mitigated and an important new research resource made available.

What about Data Quality?

Perhaps the chief concern with the prospect of web-based psychiatric research regards the quality of the data, both cognitive and self-report. The good news from psychologists is that online cognitive data closely mirrors those of lab-based studies, evidenced by the straightforward replication of numerous classic psychology effects, even those relying on reaction times (Crump et al., 2013) and more complex forms of reinforcement learning (Gillan et al., 2015b). That being said, the data are undoubtedly noisier, in large part due to the lack of environmental control. Participants are probably watching television, listening to radio, or interacting with others while performing tasks, and as one might imagine, students have been shown to follow task instructions better when tested on-campus versus at home (Ramsey et al., 2016). Increasing sample sizes can easily offset this increase in general variability, but one does need to consider forms of variation that might confound certain correlational studies. For example, the study of addiction may be particularly problematic online as the target population may be systematically more likely to be intoxicated while performing tasks at home (something that can be controlled for in a lab using blood tests).

The upside is that on anonymous platforms like Mechanical Turk, subjects have little incentive to lie (so long as their answers are not linked to their eligibility), and so one can simply ask subjects directly and try to control for these confounds statistically. Indeed, the reliability of self-report data on Mechanical Turk is quite impressive; for example, Shapiro and colleagues found that the test retest reliability of self-reported depression symptoms after 1 week was r = 0.87 (Shapiro et al., 2013). We found that less than 1% of subjects were "caught" by a question designed to identify subjects who were not reading the self-report questions (Gillan et al., 2016), and Shapiro and colleagues found that less than 3% provided inconsistent demographics across two sessions separated by 1 week, and another 3% showed evidence of malingering (Shapiro et al., 2013).

Standardization and Generalizability

One of the most tantalizing aspects of online testing is that, even though it adds a great deal of untracked heterogeneity with respect to testing conditions, the use of fully computerized testing also promotes methodological standardization and therefore reproducibility of research findings. Specifically, it eliminates the variability inherent in providing verbal task instructions, eliminates the possibility that patient groups are systematically coached or given supplemental or special instructions, and eliminates other forms of unintentional experimenter influence that don't typically show up in research papers. As a result, the data from online research studies are perhaps as generalizable as they can get.

This also applies to the use of self-report in place of diagnostic interviews, a practice that will likely become the mainstay of scalable online psychiatric research. However, this piece of this enterprise is perhaps the most controversial. Many argue that trained raters are essential to accurately determine the presence or absence of DSM disorders, a.k.a. differential diagnosis. There is nonetheless a strong argument that clinician-rated measures are in some important respects less reliable than self-report. In addition to variation that comes from the (lack of) reliability of patients' responses (e.g., test, re-test), clinicians add another source of noise—inter-rater differences in interpretation of patients' responses. Online data collection in psychiatry, certainly the most scalable and ambitious forms of it, will necessarily move away from diagnostic interviews. This will likely have a positive impact on self-report instruments, which will be refined as they become more central to our research investigations and as such come to map more closely to underlying psychological and biological phenomena.

Aside from ways that online testing improves research standardization, it also can enhance generalizability in other ways. Mechanical Turk samples are more demographically representative of the US population than those recruited from college campuses (Berinsky et al., 2012). Moreover, Internet-based research removes a major barrier to participation for many individuals for whom leaving the home and traveling to attend a university campus is simply not possible. This means that studies can recruit some of the most severely disabled patients, who are missed by in-person studies. Relatedly, it also allows access to special samples that can be difficult to recruit—for example, while individuals on Mechanical Turk are just as depressed and generally anxious as the rest of us, they have elevated levels of social anxiety (i.e., seven times greater than the 6.8% rate of 12-month prevalence in the general population) (Shapiro et al., 2013).

Conclusion

The Internet provides many new opportunities at a time of great change in psychiatry research. Size matters: "big data" is needed to overhaul classification by linking psychiatric states to their neurobiological etiology. In this way, theory-based treatment development research will see a revival. The Internet is not the only source of large datasets, of course, but it has many benefits over more traditional lab-based approaches, aside from its cost effectiveness. It will allow us to reach elusive populations in creative new ways, to conduct hybrid studies that mix field experimentation with controlled manipulation, and to study symptoms in the wild and in greater detail than ever before. This is an exciting time for psychiatry: a chance (and a challenge) to develop big and bold new ideas, the fruits of which can make a real difference in the clinic.

ACKNOWLEDGMENTS

We would like to thank Elizabeth A. Phelps, Michal Kosinski, and Robert Whelan for collaborating with us on one of the projects highlighted here (Gillan et al., 2016), which was funded by a Sir Henry Wellcome Postdoctoral Fellowship (101521/Z/12/Z) award to C.M.G.

REFERENCES

Berinsky, A., Huber, G., and Lenz, G. (2012). Polit. Anal. *20*, 351–368.

Chekroud, A.M., Zotti, R.J., Shehzad, Z., Gueorguieva, R., Johnson, M.K., Trivedi, M.H., Cannon, T.D., Krystal, J.H., and Corlett, P.R. (2016). Lancet Psychiatry *3*, 243–250.

Crump, M.J., McDonnell, J.V., and Gureckis, T.M. (2013). PLoS ONE *8*, e57410.

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., and Dolan, R.J. (2011). Neuron *69*, 1204–1215.

de Wit, S.J., Alonso, P., Schweren, L., Mataix-Cols, D., Lochner, C., Menchón, J.M., Stein, D.J., Fouche, J.P., Soriano-Mas, C., Sato, J.R., et al. (2014). Am. J. Psychiatry *171*, 340–349.

DeBattista, C., Kinrys, G., Hoffman, D., Goldstein, C., Zajecka, J., Kocsis, J., Teicher, M., Potkin, S., Preda, A., Multani, G., et al. (2011). J. Psychiatr. Res. *45*, 64–75.

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M.R., and Christensen, H. (2013). J. Med. Internet Res. *15*, e247.

Gillan, C., Robbins, T., Sahakian, B., van den Heuvel, O., and van Wingen, G. (2015a). Eur. Neuropsychopharmacol. *26*, 828–840.

Gillan, C.M., Otto, A.R., Phelps, E.A., and Daw, N.D. (2015b). Cogn. Affect. Behav. Neurosci. *15*, 523–536.

Gillan, C.M., Kosinski, M., Whelan, R., Phelps, E.A., and Daw, N.D. (2016). eLife *5*, http://dx.doi.org/10.7554/eLife.11305.

Kendler, K.S., Aggen, S.H., Knudsen, G.P., Røysamb, E., Neale, M.C., and Reichborn-Kjennerud, T. (2011). Am. J. Psychiatry *168*, 29–39.

Kessler, R.C., van Loo, H.M., Wardenaar, K.J., Bossarte, R.M., Brenner, L.A., Ebert, D.D., de Jonge, P., Nierenberg, A.A., Rosellini, A.J., Sampson, N.A., et al. (2016). Epidemiol. Psychiatr. Sci. Published online January 26, 2016.

Kosinski, M., Stillwell, D., and Graepel, T. (2013). Proc. Natl. Acad. Sci. USA *110*, 5802–5805.

Kramer, A.D., Guillory, J.E., and Hancock, J.T. (2014). Proc. Natl. Acad. Sci. USA *111*, 8788–8790.

McNab, F., Zeidman, P., Rutledge, R.B., Smittenaar, P., Brown, H.R., Adams, R.A., and Dolan, R.J. (2015). Proc. Natl. Acad. Sci. USA *112*, 6515–6518.

Otto, A.R., Fleming, S.M., and Glimcher, P.W. (2016). Psychol. Sci. *27*, 299–311.

Ramsey, S., Thompson, K., McKenzie, M., and Rosenbaum, A. (2016). Comput. Human Behav. *58*, 354–360.

Rutledge, R.B., Skandali, N., Dayan, P., and Dolan, R.J. (2014). Proc. Natl. Acad. Sci. USA *111*, 12252–12257.

Schumann, G., Loth, E., Banaschewski, T., Barbot, A., Barker, G., Büchel, C., Conrod, P.J., Dalley, J.W., Flor, H., Gallinat, J., et al.; IMAGEN consortium (2010). Mol. Psychiatry *15*, 1128–1139.

Shapiro, D., Chandler, J., and Mueller, P. (2013). Clin. Psychol. Sci. http://dx.doi.org/10.1177/2167702612469015.