JAMA Psychiatry | Special Communication

Studying Healthy Psychosislike Experiences to Improve Illness Prediction

Philip R. Corlett, PhD; Sonia Bansal, PhD; James M. Gold, PhD

IMPORTANCE Distinguishing delusions and hallucinations from unusual beliefs and experiences has proven challenging.

OBSERVATIONS The advent of neural network and generative modeling approaches to big data offers a challenge and an opportunity; healthy individuals with unusual beliefs and experiences who are not ill may raise false alarms and serve as adversarial examples to such networks.

CONCLUSIONS AND RELEVANCE Explicitly training predictive models with adversarial examples should provide clearer focus on the features most relevant to casehood, which will empower clinical research and ultimately diagnosis and treatment.

JAMA Psychiatry. 2023;80(5):515-517. doi:10.1001/jamapsychiatry.2023.0059
Published online March 8, 2023.

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Philip R. Corlett, PhD, Department of Psychiatry, Yale University School of Medicine, 34 Park St, New Haven, CT 06519 (philip.corlett@yale.edu).

It has long been appreciated that "phenomena are best understood when placed within their series, studied in their germ and in their over-ripe decay, and compared with their exaggerated and degenerated kindred."[1] Hallucinationlike experiences and delusionlike beliefs are common in individuals without psychoses. In psychiatry, an approach that incorporates this observation was pioneered by Meehl,[2] who argued for a psychosis continuum along which esoteric experiences and beliefs fell, coupled with social and biological insults that were specific to illness.

The propriety and utility of continuum approaches continues to be debated.[3,4] Under a continuum model, the symptoms of serious mental illness can be found in an attenuated form distributed on a continuum in the general population, with clinical cases at the extreme. In such frameworks, attenuated symptomlike experiences in the general population can serve as useful models for clinical symptoms. This approach has been the cornerstone of empirical efforts toward predictive processing accounts of psychotic symptoms.[5,6] Here we suggest individuals with schizotypal dispositions who experience attenuated psychotic symptoms may also provide useful foils as the field of psychiatry joins the inexorable march toward precision approaches driven by machine learning. That is, we might learn not only from what they share with people with schizophrenia, but also, and perhaps more usefully, from how they differ.

Specifically, people who experience delusionlike beliefs and hallucinations in the absence of illness may serve as adversarial examples—that is, data points that fool machine learning algorithms, such as an apparently noisy image that is classified as containing a cat when it does not. Rather than being catastrophic for the enterprise, we see these individuals as exemplars that can clarify the features of psychopathology that are pathognomonic of psychotic illness. Learning about the ways in which they are similar to, and different from, patients with psychotic illnesses will enhance the precision of predictions about diagnosis and prognosis, just as adversarial examples in computer vision can improve model performance.

Precision psychiatry strives to forecast yet unobserved outcomes. In some cases, it is based on a generative model.[7] The promise is considerable. For example, based on a simple intake questionnaire, it is possible to predict the outcome of a particular drug treatment for a specific person with depression with 64.6% accuracy.[8] However, models are only as good as the data on which they are trained. Training models on biased data will result in recapitulating those biases in the models' predictions.[9] Furthermore, computer vision (advances in which have been foundational to the broader application of machine learning) has furnished a new puzzle—namely, adversarial examples,[10] which we see as both a challenge and an opportunity and which reprise some of the difficulties of traditional diagnosis.

Computer vision can classify natural images with equivalent accuracy to human adults. Convolutional neural networks comprise multiple layers of interconnected artificial neurons, inspired by visual cortices. They have an input layer of neurons, a series of hidden layers, and an output layer. Nodes within layers and across layers can be connected. Each connection from a particular node can transmit signals to other nodes. With learning, this signal transduction increases or decreases, captured in the network as the weight of the connections between nodes. The output (for example, whether some object is present in the image) is a function of those weights.

Different layers of nodes perform different computations over the signals to achieve the network's end. For example, convolutional networks take advantage of the hierarchical patterns in data and assemble patterns of increasing complexity using smaller and simpler patterns embossed in the filters of each hidden layer. They learn to identify images that contain cats by analyzing example images that have been manually labeled as containing a cat or no cat, taking those labeled images as their input (creating activation patterns across their input nodes), making a prediction based on the subsequent responses in their hidden layers, and outputting that

prediction. Any mismatch between predicted and actual label (or prediction error) is used to update the weights so that, with experience, the network gets better and better at identifying cats. This technology has been adopted to screen luggage at the airport and perform diagnoses on radiological scans.

Recently, the speed and precision of training has been improved by connecting 2 networks and having them train one another. These generative adversarial networks have 3 key components: a training set of images (containing, for example, cats), a generator network, and a critic network.

First, the generator is trained with the training set ("this is a picture of a cat"). Then it is inverted to produce novel content (images of cats). The critic decides whether a generated image belongs among the training set ("is this a cat?"). It is reinforced for correctly rejecting what the generator produces. The generator is reinforced for fooling the critic. Thus, the 2 networks train one another.

However, generative adversarial networks often err in bizarre ways. Two of its vulnerabilities are fooling images and perturbed images produced by the generator, both of which are misclassified by the critic. Fooling images bear no ostensible resemblance to the category to which they are attributed by the networks. For example, an image of noisy black and white dots like static on a TV may be classified as containing a cat when nothing approximating a feline appears to be present. Perturbed images appear to straightforwardly contain the category (cat) but are grossly miscategorized (eg, as having no cat or, in more complex networks, as having a dog or even elephant) due to some apparently imperceptible variance. We would like to suggest that people who believe in conspiracy theories or new religious movements and people who hear voices but do not have clinical diagnoses might serve as adversarial examples to a network trained on the phenomenal features and cognitive and neural mechanisms of psychosis that we have accrued as a field so far. Such individuals might be classified as having a psychotic illness when in fact they do not.

In foundational work on the psychosis continuum, Garety and colleagues[11] created the Peters Delusions Inventory, a self-report questionnaire that asks whether participants endorse a series of delusionlike beliefs, the contents of which were drawn from existing knowledge and scales which capture the themes of delusions. The questions are phrased in such a manner that they might detect subtle or minimal versions of the belief. For example, they use relative *as if* statements. To capture healthy beliefs that resemble thought insertion, they ask, "Do your thoughts ever feel alien to you in some way?" If a belief is endorsed, further questions are posed: "How convinced are you that it is true?", "How often do you think about it?", and "How distressing do you find it?" People without a psychotic illness endorse a mean (SD) of 7 (4) unusual beliefs compared to 12 (6) in patients with schizophrenia (clearly, these ranges overlap). Patients are typically more convinced, preoccupied, and distressed than nonpatient control participants. However, in studies of participants in new religious movements (eg, druidry and spiritual healing), it is possible to find people who do not meet diagnostic criteria for schizophrenia but who endorse as many unusual beliefs as patients with schizophrenia and who are as convinced by and preoccupied with these beliefs.[12] Importantly, such individuals are not as distressed by their delusionlike beliefs as people with schizophrenia.[13] The same may be true of individuals who endorse multiple conspiracy theories. Nevertheless, when people without psychosis are distressed by the delusions they endorse, their task-derived prediction-error brain response, although lower, begins to resemble that of people with schizophrenia[14] and people who have been administered ketamine.[15] Apparently, delusion-related distress, rather than the presence of delusions per se, would be more pathognomonic of psychotic illness. The same may be true of auditory verbal hallucinations.

Voice hearing can also be observed among people who are neither unwell nor have psychosis. This, too, appears to be distributed on a continuum. People often report hearing a deceased loved one speak to them, particularly proximal to a bereavement experience. Religious practitioners endorse hearing God speak to them on occasion. Further along the continuum, people who identify as clairaudient psychics report hearing voices daily without meeting illness criteria.[16] The voices have the same loudness, syntax, and semantics, as those reported by people with schizophrenia.[16] However, they are rarely as distressing, and clairaudient participants report a greater degree of volitional control over the onset and offset of voices.[16]

It may be that hallucinations and delusions are merely accessory symptoms and that other features better portend illness, such as thought disorder, trauma, or dissociation. However, individuals who hear voices but do not have clinical diagnoses also evince considerable trauma, dissociation,[17] and thought disorder,[18] although there is room for more careful phenomenological dissection of the clinical and nonclinical experiences.[19] In terms of morphology, there is some evidence that shorter length of the paracingulate sulcus is associated with the presence of hallucinations in schizophrenia.[20] Some groups claim that the sulcus is not shortened in those who hear voices but lack clinical diagnoses. This would mean that paracingulate sulcus length could be used to signify or even portend auditory verbal hallucinations that are clinically actionable.[21] However, when we measured the length of the sulcus in people who claim to be clairaudient psychics (who, we contend, are nearer to people with schizophrenia along the continuum) it was shorter.[22] This suggests that any generative model giving weight to paracingulate sulcus length in predicting psychotic illness would raise false alarms with respect to clairaudient psychic cases. Such errors seem catastrophic for the machine learning enterprise in psychiatry, but there may be a way forward.

First, humans appear to have a remarkable ability to identify adversarial images where convolutional networks are prone to raise false alarms. Confronted with black and white static images that a network might mistake for a cat, humans correctly guess which image will be misclassified as a cat.[10] This may reveal which features are engendering the errors. Next, humans can tell the difference between patients with psychosis and individuals who hear voices but do not have clinical diagnoses. This ability is not limited to trained clinicians. Anthropologists have shown that members of the lay public can discern between people with mental illness and shamans, despite ostensible similarities between their experiences and beliefs.[23] By testing and training neural networks on adversarial examples and supervising their false alarms, might we leverage the opportunities of adversarial examples, such as individuals who ascribe to psychic practices and conspiracy theories? In this way, we can hone in on the important features that portend illness, disability, and distress, rather than features that are distributed more widely, even in individuals whose experiences closely resemble those of people with schizophrenia. We further suggest

that healthy individuals without psychosis and with attenuated but nonclinical psychotic symptomlike experiences might serve as adversarial examples on which network model aiming to predict schizophrenia could be trained, to appropriately weight the variables entering into the prediction of illness. These might include people in clinical high-risk studies who do not quite meet the threshold for psychosis risk state but who nevertheless evince some degree of distress and impairment.[24] Any differential prediction of conversion[25] might impact the ongoing debate regarding the risk state and whether it merely represents an earlier phase of the illness course. Those with attenuated psychotic symptoms who do not convert might also serve this adversarial function. This training may come in the form of human supervision, correcting and reorienting the networks. It may come in terms of basic empirical science, constraining which features enter into predictive models, removing those that have been shown to feature significantly in people who assert psychic ability or are prone to believe in conspiracy theories. Or the training could be more direct. There are training regimens for neural networks that focus on the features and underlying distributions of adversarial examples.[26] In computer vision, these regimens can greatly improve performance. Some combination of these ought to benefit precision approaches in psychiatry, helping them make better predictions and leveraging the considerable expertise of human clinicians and scientists.

ARTICLE INFORMATION

Accepted for Publication: January 9, 2023.

Published Online: March 8, 2023. doi:10.1001/jamapsychiatry.2023.0059

Author Affiliations: Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut (Corlett); Wu Tsai Institute, Yale University, New Haven, Connecticut (Corlett); Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland School of Medicine, Baltimore (Bansal, Gold).

Conflict of Interest Disclosures: Dr Corlett is cofounder and equity holder of Tetricus Labs, a startup incorporated in 2022 that attempts to diagnose and treat mental illnesses using behavioral testing and artificial neural networks. Dr Gold reported personal fees from Vera Sci outside the submitted work. No other disclosures were reported.

Funding/Support: Drs Corlett, Bansal, and Gold were supported by National Institute of Mental Health grants R01MH112887 (all), R01 MH120090 (Dr Gold), and R01 MH120089 (Dr Corlett). Dr Corlett was also supported in part by the State of Connecticut, Department of Mental Health and Addiction Services.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclaimer: This publication does not express the views of the Department of Mental Health and Addiction Services or the State of Connecticut. The views and opinions expressed are those of the authors.

REFERENCES

1. James W. *Varieties of Religious Experience: A Study in Human Nature, Being the Gifford Lectures on Natural Religion Delivered at Edinburgh in 1901-1902*. Longmans, Green & Co; 1917.

2. Meehl PE. Schizotaxia, schizotypy, schizophrenia. *Am Psychol*. 1962;17(12):827-838. doi:10.1037/h0041029

3. David AS. Why we need more debate on whether psychotic symptoms lie on a continuum with normality. *Psychol Med*. 2010;40(12):1935-1942. doi:10.1017/S0033291710000188

4. Lawrie SM. Whether "psychosis" is best conceptualized as a continuum or in categories is an empirical, practical and political question. *World Psychiatry*. 2016;15(2):125-126. doi:10.1002/wps.20325

5. Schmack K, Gòmez-Carrillo de Castro A, Rothkirch M, et al. Delusions and the role of beliefs in perceptual inference. *J Neurosci*. 2013;33(34):13701-13712. doi:10.1523/JNEUROSCI.1778-13.2013

6. Schmack K, Bosc M, Ott T, Sturgill JF, Kepecs A. Striatal dopamine mediates hallucination-like perception in mice. *Science*. 2021;372(6537):eabf4740. doi:10.1126/science.abf4740

7. Bzdok D, Varoquaux G, Steyerberg EW. Prediction, not association, paves the road to precision medicine. *JAMA Psychiatry*. 2021;78(2):127-128. doi:10.1001/jamapsychiatry.2020.2549

8. Chekroud AM, Zotti RJ, Shehzad Z, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry*. 2016;3(3):243-250. doi:10.1016/S2215-0366(15)00471-X

9. Birhane A. Algorithmic injustice: a relational ethics approach. *Patterns (N Y)*. 2021;2(2):100205. doi:10.1016/j.patter.2021.100205

10. Zhou Z, Firestone C. Humans can decipher adversarial images. *Nat Commun*. 2019;10(1):1334. doi:10.1038/s41467-019-08931-6

11. Peters ER, Joseph SA, Garety PA. Measurement of delusional ideation in the normal population: introducing the PDI (Peters et al Delusions Inventory). *Schizophr Bull*. 1999;25(3):553-576. doi:10.1093/oxfordjournals.schbul.a033401

12. Smith L, Riley S, Peters ER. Schizotypy, delusional ideation and well-being in an American new religious movement population. *Clin Psychol Psychother*. 2009;16(6):479-484. doi:10.1002/cpp.645

13. Peters E, Day S, McKenna J, Orbach G. Delusional ideation in religious and psychotic populations. *Br J Clin Psychol*. 1999;38(1):83-96. doi:10.1348/014466599162683

14. Corlett PR, Fletcher PC. The neurobiology of schizotypy: fronto-striatal prediction error signal correlates with delusion-like beliefs in healthy people. *Neuropsychologia*. 2012;50(14):3612-3620. doi:10.1016/j.neuropsychologia.2012.09.045

15. Corlett PR, Honey GD, Aitken MR, et al. Frontal responses during learning predict vulnerability to the psychotogenic effects of ketamine: linking cognition, brain activity, and psychosis. *Arch Gen Psychiatry*. 2006;63(6):611-621. doi:10.1001/archpsyc.63.6.611

16. Powers AR III, Kelley MS, Corlett PR. Varieties of voice-hearing: psychics and the psychosis continuum. *Schizophr Bull*. 2017;43(1):84-98. doi:10.1093/schbul/sbw133

17. Baumeister D, Sedgwick O, Howes O, Peters E. Auditory verbal hallucinations and continuum models of psychosis: a systematic review of the healthy voice-hearer literature. *Clin Psychol Rev*. 2017;51:125-141. doi:10.1016/j.cpr.2016.10.010

18. Sommer IE, Derwort AM, Daalman K, de Weijer AD, Liddle PF, Boks MP. Formal thought disorder in non-clinical individuals with auditory verbal hallucinations. *Schizophr Res*. 2010;118(1-3):140-145. doi:10.1016/j.schres.2010.01.024

19. Luhrmann TM, Alderson-Day B, Bell V, et al. Beyond trauma: a multiple pathways approach to auditory hallucinations in clinical and nonclinical populations. *Schizophr Bull*. 2019;45(45)(suppl 1):S24-S31. doi:10.1093/schbul/sby110

20. Garrison JR, Fernyhough C, McCarthy-Jones S, Haggard M, Simons JS; Australian Schizophrenia Research Bank. Paracingulate sulcus morphology is associated with hallucinations in the human brain. *Nat Commun*. 2015;6:8956. doi:10.1038/ncomms9956

21. Garrison JR, Fernyhough C, McCarthy-Jones S, Simons JS, Sommer IEC. Paracingulate sulcus morphology and hallucinations in clinical and nonclinical groups. *Schizophr Bull*. 2019;45(4):733-741. doi:10.1093/schbul/sby157

22. Powers AR, van Dyck LI, Garrison JR, Corlett PR. Paracingulate sulcus length is shorter in voice-hearers regardless of need for care. *Schizophr Bull*. 2020;46(6):1520-1523. doi:10.1093/schbul/sbaa067

23. Murphy JM. Psychiatric labeling in cross-cultural perspective. *Science*. 1976;191(4231):1019-1028. doi:10.1126/science.1251213

24. Fusar-Poli P, Rocchetti M, Sardella A, et al. Disorder, not just state of risk: meta-analysis of functioning and quality of life in people at high risk of psychosis. *Br J Psychiatry*. 2015;207(3):198-206. doi:10.1192/bjp.bp.114.157115

25. Gold JM, Corlett PR, Strauss GP, et al. Enhancing Psychosis Risk Prediction Through Computational Cognitive Neuroscience. *Schizophr Bull*. 2020;46(6):1346-1352. doi:10.1093/schbul/sbaa091

26. Xie C, Tan M, Gong B, Wang J, Yuille A, Le QV. Adversarial examples improve image recognition. 2020 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:819-828. doi:10.1109/cvpr42600.2020.00090