large dimensionality allows them to directly verify ETH predictions experimentally. Specifically, Kaufman *et al.* prepare two copies of the same system, with exactly one boson on every site. After a quantum quench, which allows particles to hop, correlations grow and the system becomes entangled. By performing a many-body interference experiment on the two copies, as suggested in (*9*) and tested experimentally in (*10*), the entanglement entropy of different subsystems as well as the entropy of the full state was measured (see the figure). Although the system as a whole remains pure, small subsystems are found to become mixed after a short transient time. Indeed, the reduced density matrices of one- and two-site subsystems become indistinguishable from those of a thermal ensemble. This equivalence is verified by direct observation of the particle occupation distribution and by comparing it with the equilibrium predictions. A recent experiment in a smaller system of three superconducting qubits (*11*) verified that the full time-averaged density matrix becomes thermal in chaotic regimes; another direct consequence of ETH (*8*).

Not only does ETH validate the use of statistical mechanics; there are also many important implications of these ideas to future science and technology. Understanding the microscopic structure of complex systems can provide the necessary tools and intuition for designing systems with similar or better performance than those found in nature, which often operate efficiently in far from ideal conditions. Understanding the conditions leading to the breakdown of ETH could be important for developing new technologies not suffering from the usual thermodynamic limitations. Remarkably, what first appeared to be an issue of controversy in quantum mechanics has provided an elegant solution to the problem of thermalization. It is the existence of individual highly entangled eigenstates that allows the somewhat ambiguous coarse-graining required in standard classical arguments to be dropped. Interestingly, ETH can be applied to systems near the classical limit, providing a simple mathematical framework to understand unanswered questions in classical chaotic systems. ■

REFERENCES

1. A. M. Kaufman *et al.*, *Science* **353**, 794 (2016).
2. M. Rigol, V. Dunjko, M. Olshanii, *Nature* **452**, 854 (2008).
3. J. Eisert, M. Friesdorf, C. Gogolin, *Nat. Phys.* **11**, 124 (2015).
4. S. Goldstein *et al.*, *Proc. R. Soc. A* **466**, 3203 (2010).
5. F. Borgonovi *et al.*, *Phys. Rep.* **626**, 1 (2016).
6. J. M. Deutsch, *Phys. Rev. A* **43**, 2046 (1991).
7. M. Srednicki, *Phys. Rev. E* **50**, 888 (1994).
8. L. D'Alessio *et al.*, *Adv. Phys.* **65**, 239 (2016).
9. A. J. Daley *et al.*, *Phys. Rev. Lett.* **109**, 020505 (2012).
10. R. Islam *et al.*, *Nature* **528**, 77 (2015).
11. C. Neill *et al.*; https://arxiv.org/abs/1601.00600 (2016).
12. M. Schreiber *et al.*, *Science* **349**, 842 (2015).
13. J.-Y. Choi *et al.*, *Science* **352**, 1547 (2016).

10.1126/science.aah5776

ECONOMICS

Fighting poverty with data

Machine learning algorithms measure and target poverty

By **Joshua Evan Blumenstock**

Policy-makers in the world's poorest countries are often forced to make decisions based on limited data. Consider Angola, which recently conducted its first postcolonial census. In the 44 years that elapsed between the prior census and the recent one, the country's population grew from 5.6 million to 24.3 million, and the country experienced a protracted civil war that displaced millions of citizens. In situations where reliable survey data are missing or out of date, a novel line of research offers promising alternatives. On page 790 of this issue, Jean *et al.* (*1*) apply recent advances in machine learning to high-resolution satellite imagery to accurately measure regional poverty in Africa.

Traditionally, wealth and poverty are measured through surveys of household income and consumption (*2*). These data provide a critical input to the world's most prominent antipoverty programs, from basic cash transfer programs to multifaceted aid programs designed to target the extreme poor (*3*). However, nationally representative surveys cost tens to hundreds of millions of dollars to collect, and many developing countries go for decades without updating their estimates.

"...there is exciting potential for adapting machine learning to fight poverty."

Over the past few decades, researchers have begun to develop different techniques for estimating poverty remotely. Initial work explored the potential of "nightlights" data: satellite photographs taken at night that capture light emitted from Earth's surface. Since such imagery first became available in the early 1970s, it was evident that wealthy regions tended to shine brightest (*4*). Recent studies have found a strong correlation between nightlight luminosity and traditional measures of economic productivity and growth (*5*, *6*). Nightlight-based measures are now frequently used by researchers, for instance to study the impact of sanctions on the economy of North Korea (*7*), where official statistics are dubious.

A series of studies in wealthy nations explore how data from the internet and social media can provide proxies for economic activity (*8*, *9*). Mining the tweets and search queries of millions of individuals promises real-time alternatives to more traditional methods of data collection. However, these approaches are less relevant to remote and developing regions, where internet infrastructure is limited and few people use social media.

In developing countries, researchers have found ways to measure wealth and poverty using the digital footprints left behind in the transaction logs of mobile phones, which are increasingly ubiquitous even in very poor regions. Regional patterns of mobile phone use correlate with the regional distribution of wealth (*10*). This relationship persists at the individual level, such that machine learning algorithms can infer an individual subscriber's socioeconomic status directly from his or her history of mobile phone use. The individual predictions can be aggregated into regional measures of wealth that are about as accurate as a 5-year-old household survey (*11*). Phone-based proxies for wealth are beginning to be used in research, e.g., to understand how new technologies differentially benefit the wealthy and the poor (*12*) and to assess the creditworthiness of would-be borrowers (*13*).

Although promising, these nontraditional methods have caveats. As Jean *et al.* show, nightlights data are less effective at differentiating between regions at the bottom end of the income distribution, where satellite images appear uniformly dark. And mobile phone data are owned by mobile phone operators and are generally not available to policy-makers. By contrast, Jean *et al.* use only publicly available data.

Taking nightlights as their starting point, the authors have devised a clever technique to also extract information from daytime satellite imagery. Daytime imagery is taken at much higher resolution than nighttime imagery. It thus contains visible features—such as paved roads and metal roofs—that make it possible to differentiate between poor and ultrapoor regions. Jean *et al.*'s insight was to apply state-of-the-art deep learning algorithms to the daytime imagery to extract these features. When given large quantities of data with labeled patterns, these algorithms

School of Information, University of California, Berkeley, CA 94720, USA. Email: jblumenstock@berkeley.edu

excel at generalizing those patterns to new data. For instance, search engines use this technology to automatically label the contents of billions of internet photos.

The authors use a convolutional neural network to learn the relationship between millions of daytime satellite images (which are rich in detail) and nighttime images (where light areas are assumed to be wealthy). In this way, the network learns which features in the daytime imagery are indicative of economic activity (see the figure). Knowledge of those features enabled the authors to accurately reconstruct survey-based indicators of regional poverty, improving on results from simpler models that relied solely on nightlights or mobile phone data.

How might these results change the way that we measure and target poverty? Perhaps the most immediate application is as a source of inexpensive, interim national statistics. Jean *et al.*'s results indicate that a model trained in one country can be used in another, creating options for countries where no recent survey data exist. For social welfare programs, some of which already use satellite imagery to identify eligible recipients (*14*), higher-fidelity estimates of poverty can help to ensure that resources get to those with the greatest need.

Other applications are on the horizon. Remotely sourced satellite and mobile phone data are updated frequently and can be used to generate nearly real-time estimates of regional vulnerability. Once it is possible to estimate short-term changes in wealth and poverty, new approaches to program monitoring and impact evaluation will follow.

Considerable validation and calibration are required before proof-of-concept studies such as that of Jean *et al.* can be used in practice. However, as their study illustrates, there is exciting potential for adapting machine learning to fight poverty. As the economist Sendhil Mullainathan has asked, "Why should the financial services industry, where mere dollars are at stake, be using more advanced technologies than the aid industry, where human life is at stake" (*15*)? ■

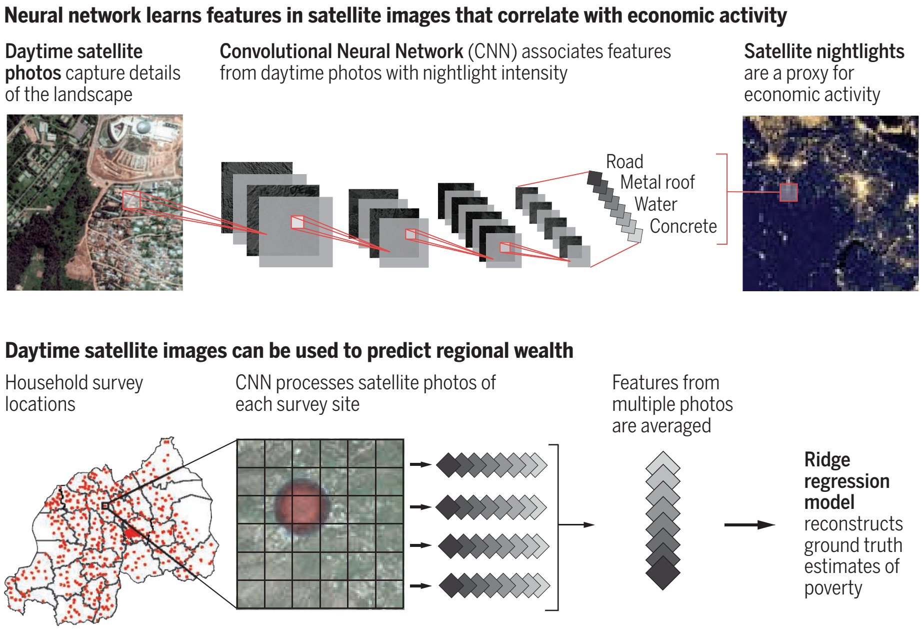
REFERENCES

1. N. Jean *et al.*, *Science* **353**, 790 (2016).
2. A. Deaton, *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy* (World Bank Publications, 1997).
3. A. Banerjee *et al.*, *Science* **348**, 1260799 (2015).
4. D. Donaldson, A. Storeygard, "Big Grids: Applications of Remote Sensing in Economics," working paper, 2016; https://dl.dropboxusercontent.com/u/2734209/Donaldson_Storeygard_JEP.pdf.
5. J. V. Henderson *et al.*, *Am. Econ. Rev.* **102**, 994 (2012).
6. X. Chen, W. D. Nordhaus, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8589 (2011).
7. Y. S. Lee, "International Isolation and Regional Inequality: Evidence from Sanctions on North Korea," working paper, 2016; https://web.stanford.edu/~yongslee/NKsanctions_030216.pdf.
8. H. Choi, H. Varian, *Econ. Rec.* **88**, 2 (2012).
9. A. Llorente *et al.*, *PLOS ONE* **10**, e0128692 (2015).
10. N. Eagle *et al.*, *Science* **328**, 1029 (2010).
11. J. E. Blumenstock *et al.*, *Science* **350**, 1073 (2015).
12. J. E. Blumenstock *et al.*, *J. Dev. Econ.* **120**, 157 (2016).
13. D. Björkegren, D. Grissen, "Behavior Revealed in Mobile Phone Usage Predicts Loan Repayment," working paper, 2015; http://dan.bjorkegren.com/files/danbjork_creditscoring.pdf.
14. J. Haushofer, J. Shapiro, *Q. J. Econ.* 10.1093/qje/qjw025 (2016).
15. S. Mullainathan, "Satellite images can pinpoint poverty where surveys can't," *New York Times*, 1 April 2016; www.nytimes.com/2016/04/03/upshot/satellite-images-can-pinpoint-poverty-where-surveys-cant.html?_r=0.

10.1126/science.aah5217

Predicting poverty

Satellite images can be used to estimate wealth in remote regions.



GRAPHIC: IMAGES COURTESY OF GOOGLE EARTH, NASA, AND J. E. BLUMENSTOCK; ADAPTED BY N.CARY/*SCIENCE*

CRYSTALLOGRAPHY

Now you see me too

Attaching chiral molecules to a chiral framework allows their molecular structures to be determined

By **Lars Öhrström**

Knowledge of three-dimensional (3D) molecular structures is crucial for scientific advances in fields ranging from materials chemistry to medicine. For solar cell materials, human proteins, or new drugs, the revelation of the exact arrangement of atoms and bonds vastly advances understanding of their properties. On page 808 of this issue, Lee *et al.* (*1*) report an approach that allows better structural data to be obtained for large, complex organic molecules that are difficult to crystallize on their own.

The method of choice to obtain structure information is single-crystal x-ray diffraction, a method so important that UNESCO declared 2014 the International Year of Crystallography. However, this method requires not only a pure substance, but also the ability to grow crystals of it—no crystals, no crystal structure data. The main complementary method, nuclear magnetic resonance, mainly provides structures of compounds in solution, often at great detail, but sometimes with inherent uncertainty, especially for chiral (handed) molecules with complicated stereochemistry.

Although long hours in the lab may produce crystals, some substances are notoriously difficult to crystallize or yield crystals with defects and disorder that prevent a complete structure determination. On the other hand, the molecular structures of small solvent molecules, trapped between the larger molecules that are the principal constituents of a specific crystal, are determined over and over again; for example, 1989 molecular structures of pyridine, C_5H_5N, are reported in the Cambridge Crystallographic Database (*2*). This occurs because the form and intermolecular interactions of the larger molecules sometimes generate voids in the crystal. Scientists

Department of Chemistry and Chemical Engineering, Chalmers University of Technology, Gothenburg, Sweden. Email: ohrstrom@chalmers.se

Fighting poverty with data

Joshua Evan Blumenstock (August 18, 2016)

Science **353** (6301), 753-754. [doi: 10.1126/science.aah5217]

Editor's Summary