

Semantics derived automatically from language corpora contain human-like biases

Aylin Caliskan, Joanna J. Bryson and Arvind Narayanan

Science **356** (6334), 183-186.
DOI: 10.1126/science.aal4230

Machines learn what people know implicitly

AlphaGo has demonstrated that a machine can learn how to do things that people spend many years of concentrated study learning, and it can rapidly learn how to do them better than any human can. Caliskan *et al.* now show that machines can learn word associations from written texts and that these associations mirror those learned by humans, as measured by the Implicit Association Test (IAT) (see the Perspective by Greenwald). Why does this matter? Because the IAT has predictive value in uncovering the association between concepts, such as pleasantness and f lowers or unpleasantness and insects. It can also tease out attitudes and beliefs—for example, associations between female names and family or male names and career. Such biases may not be expressed explicitly, yet they can prove influential in behavior.

Science, this issue p. 183; see also p. 133

ARTICLE TOOLS

<http://science.sciencemag.org/content/356/6334/183>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2017/04/12/356.6334.183.DC1>

RELATED CONTENT

<http://science.sciencemag.org/content/sci/356/6334/133.full>

REFERENCES

This article cites 17 articles, 1 of which you can access for free
<http://science.sciencemag.org/content/356/6334/183#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2017, American Association for the Advancement of Science