

Randomized & natural experiments

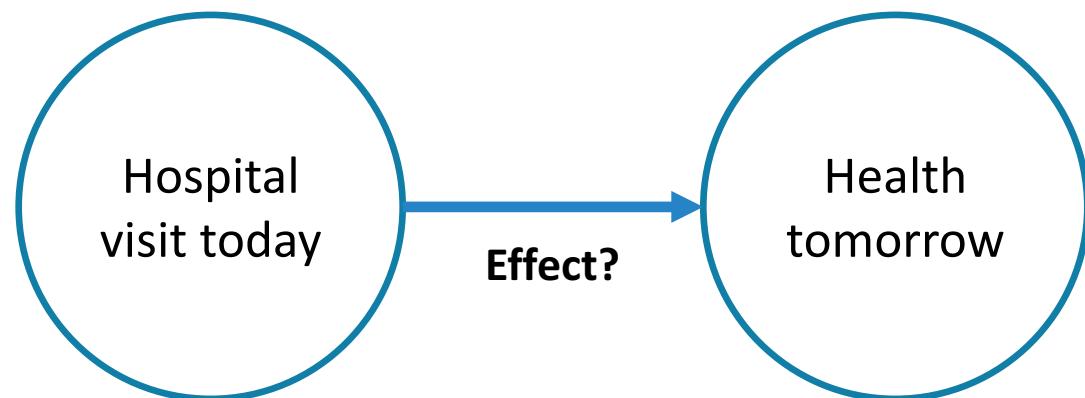
MODELING SOCIAL DATA

JAKE HOFMAN

COLUMBIA UNIVERSITY

Example: Hospitalization on health

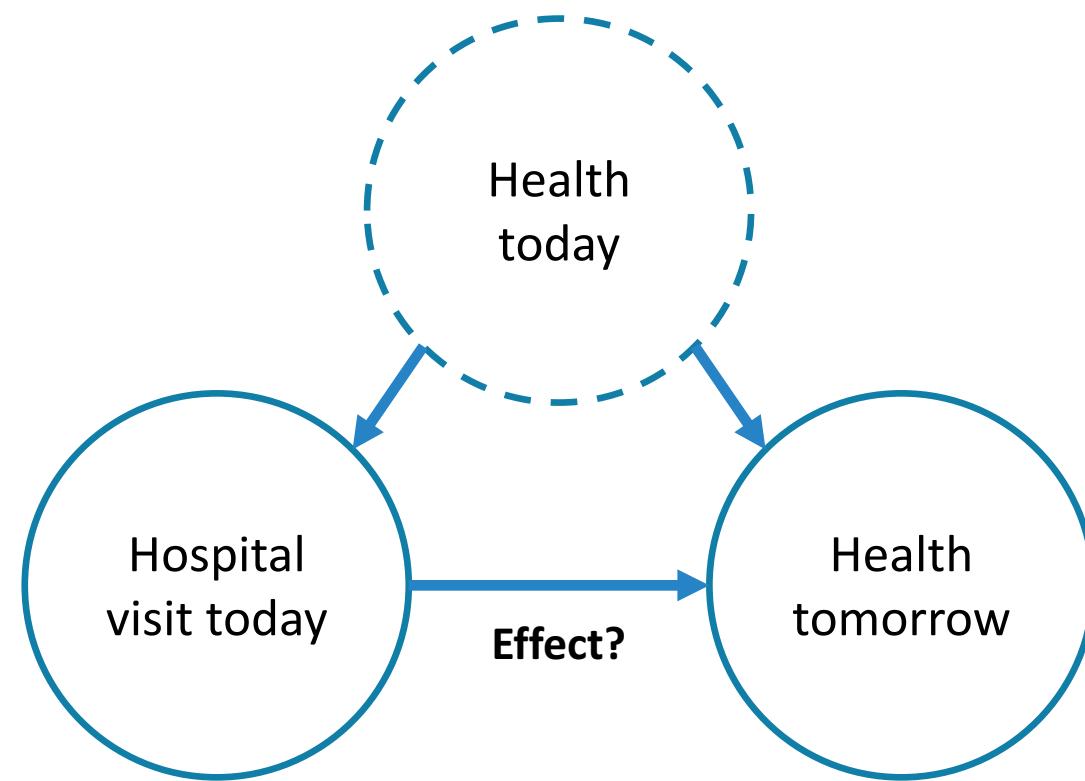
What's wrong with estimating this model from observational data?



Arrow means “X causes Y”

Confounders

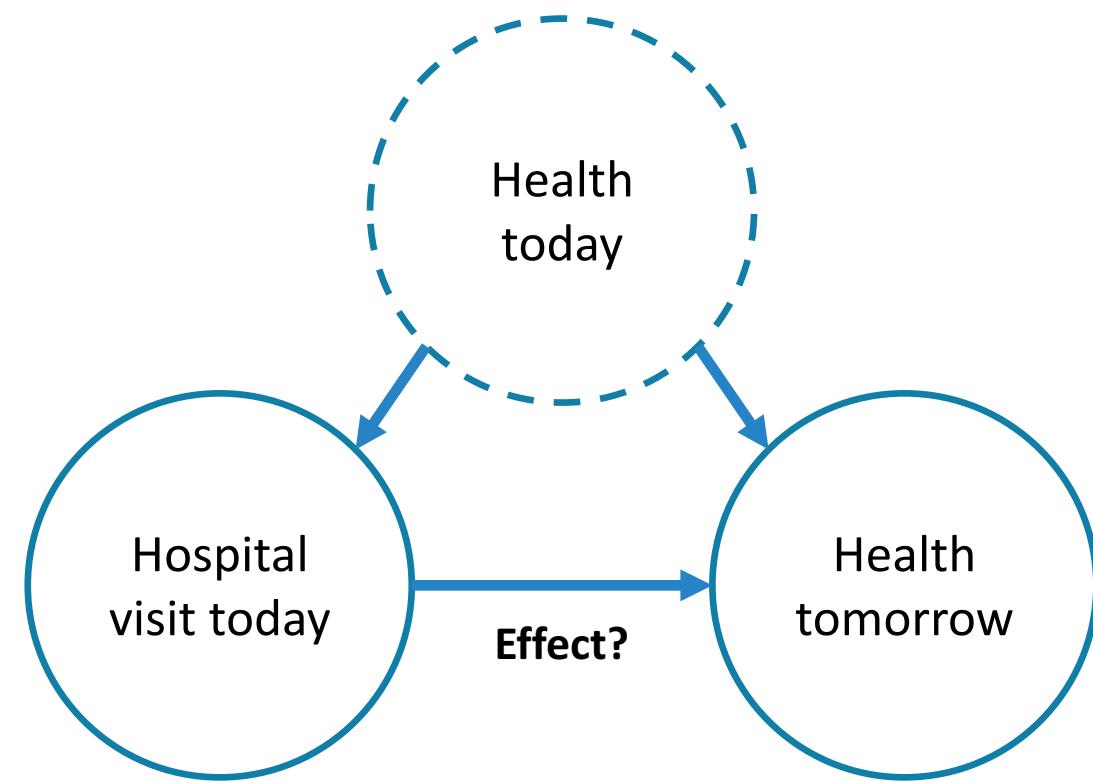
The effect and cause might be *confounded* by a common cause, and be *changing together* as a result



Dashed circle means “unobserved”

Confounders

If we *only get to observe them changing together*, we can't estimate the effect of hospitalization changing alone

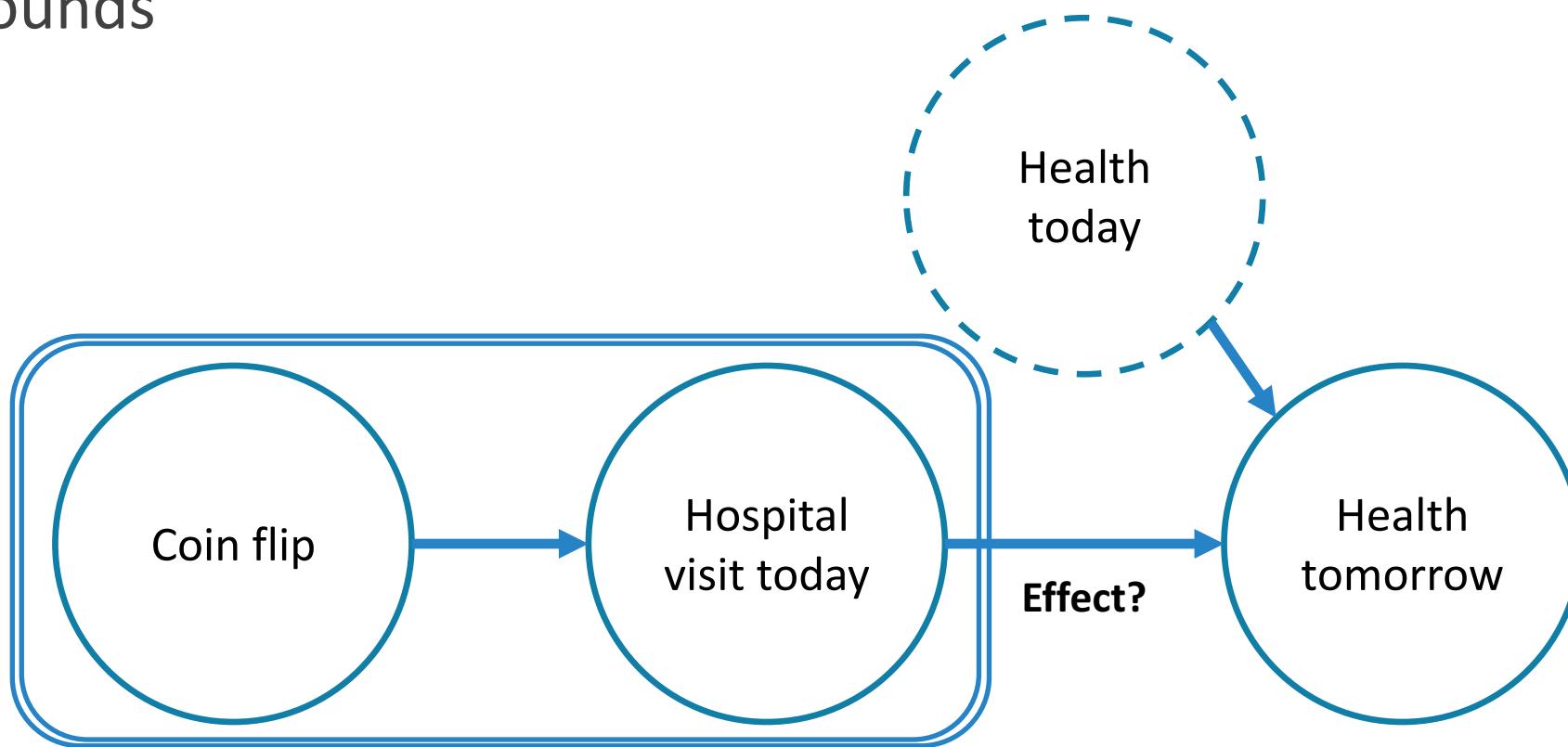


“To find out what happens when you change something, it is necessary to change it.”

-GEORGE BOX

Random assignment

Random assignment determines the treatment independent of any confounds

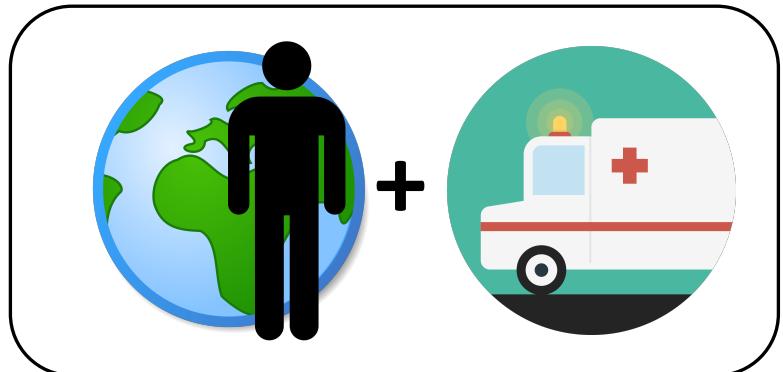


Double lines mean
“intervention”

Counterfactuals

To isolate the causal effect, we have to *change one and only one thing* (hospital visits), and compare outcomes

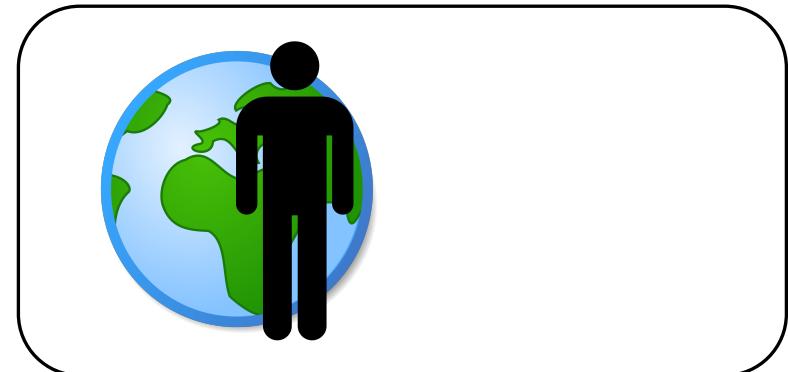
Reality



(what happened)

vs

Counterfactual

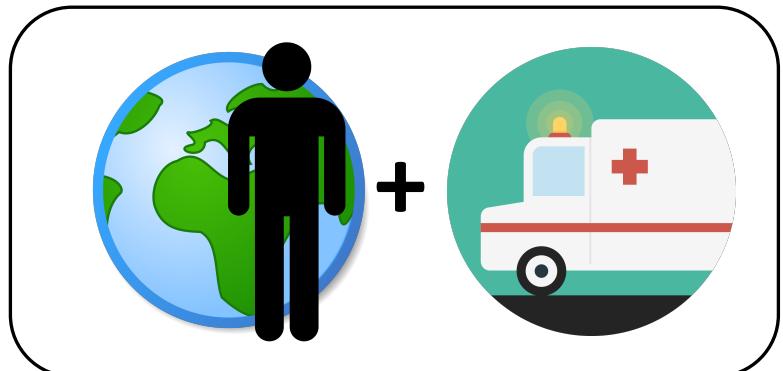


(what would have happened)

Counterfactuals

We never get to observe *what would have happened if we did something else*, so we have to estimate it

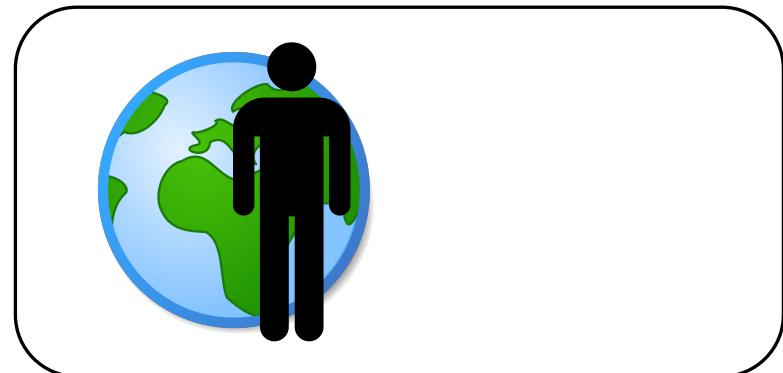
Reality



(what happened)

vs

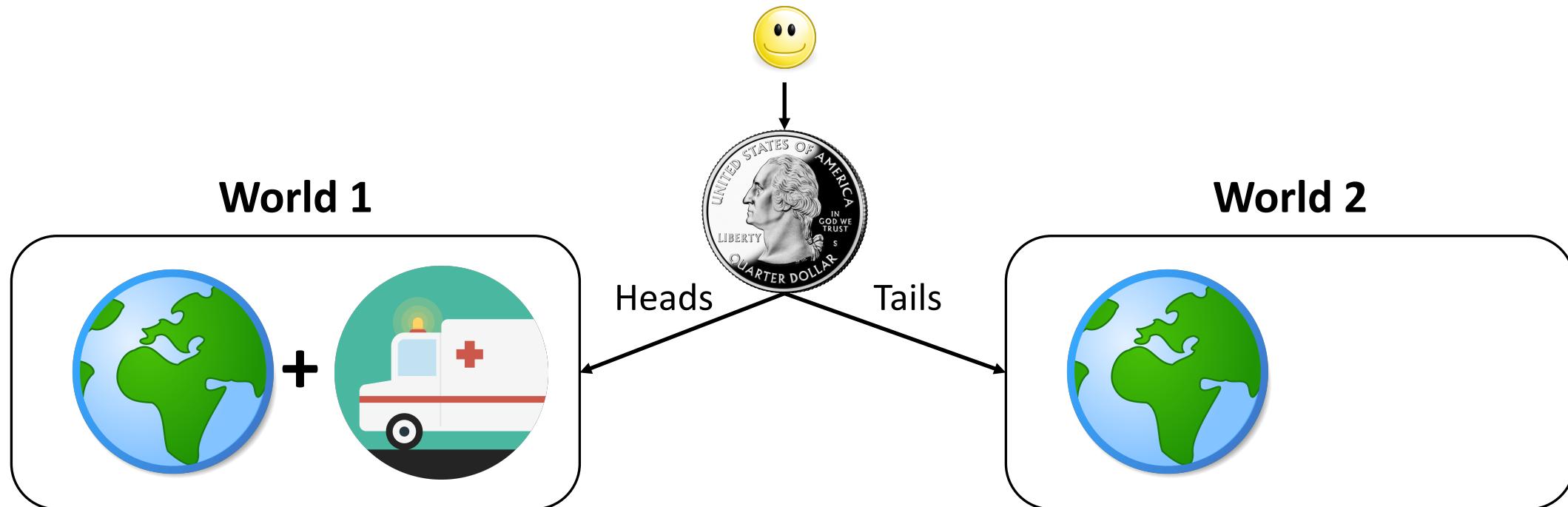
Counterfactual



(what would have happened)

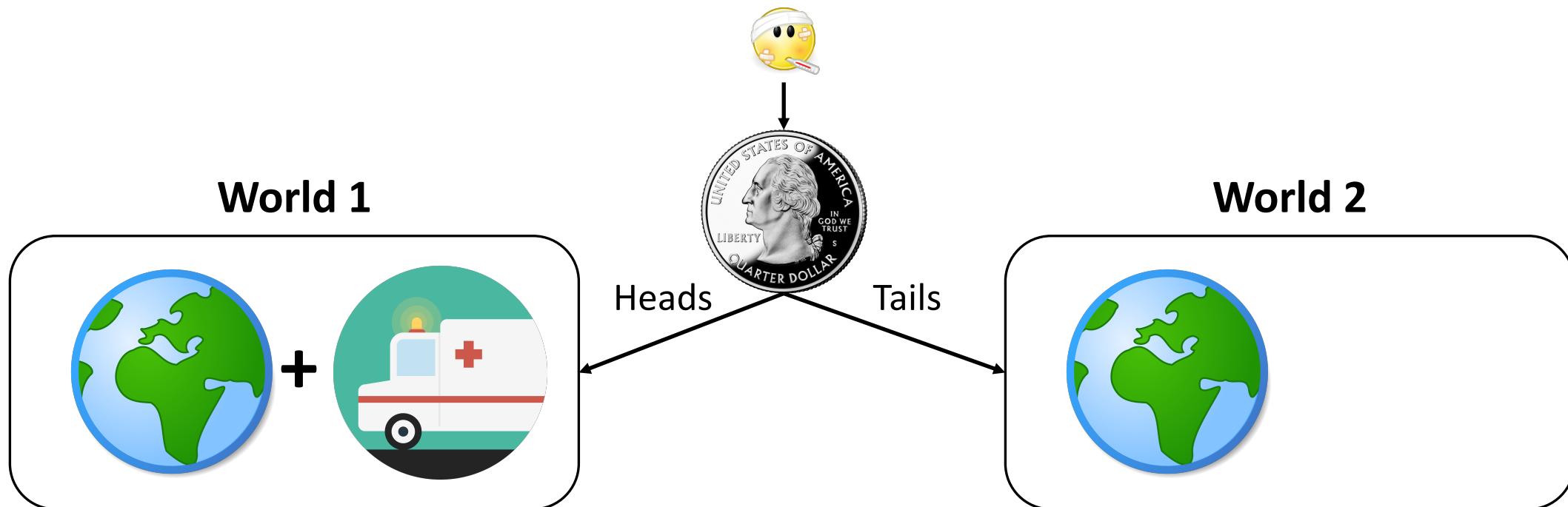
Random assignment

We can use randomization to create two groups that differ only in which treatment they receive, restoring symmetry



Random assignment

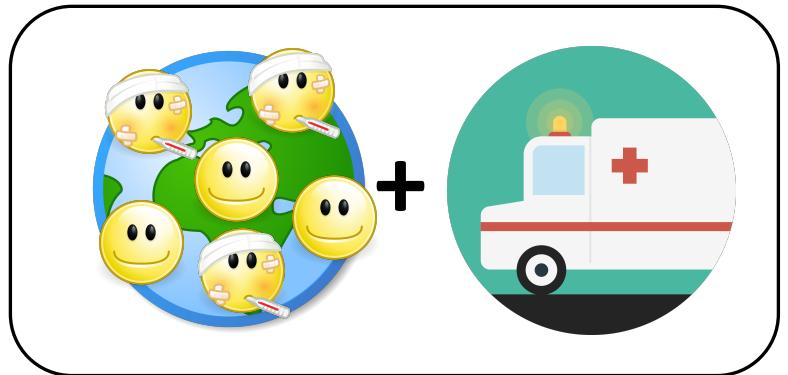
We can use randomization to create two groups that differ only in which treatment they receive, restoring symmetry



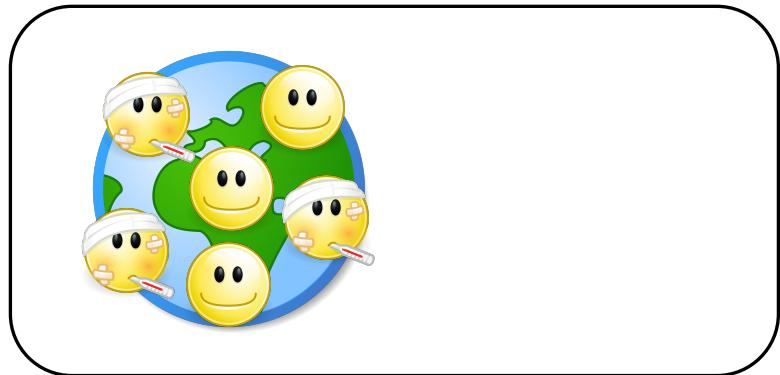
Random assignment

We can use randomization to create two groups that differ only in which treatment they receive, restoring symmetry

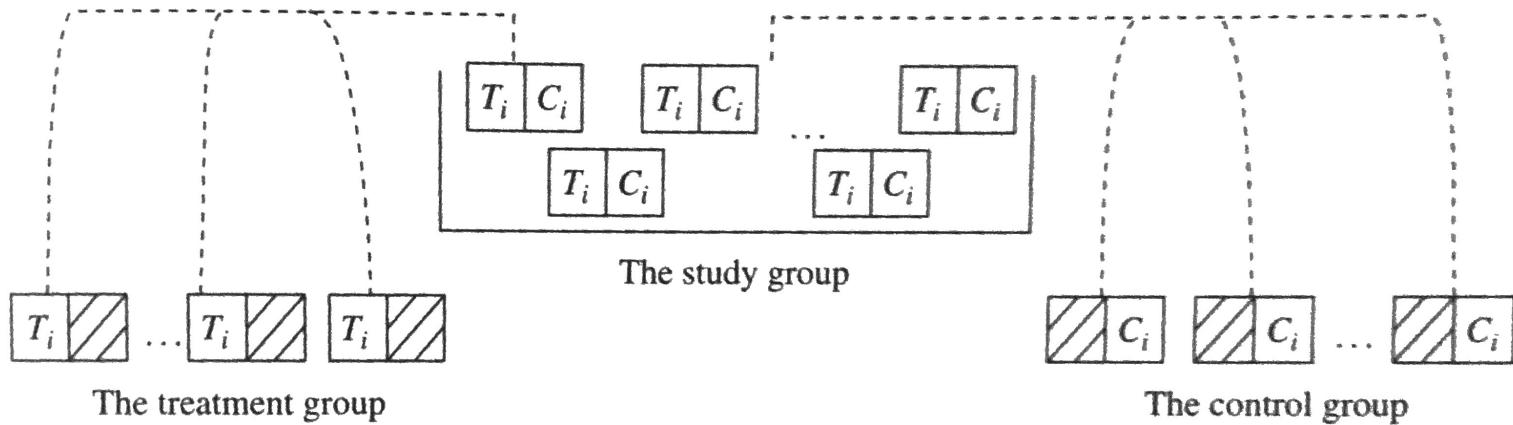
World 1



World 2



Random assignment



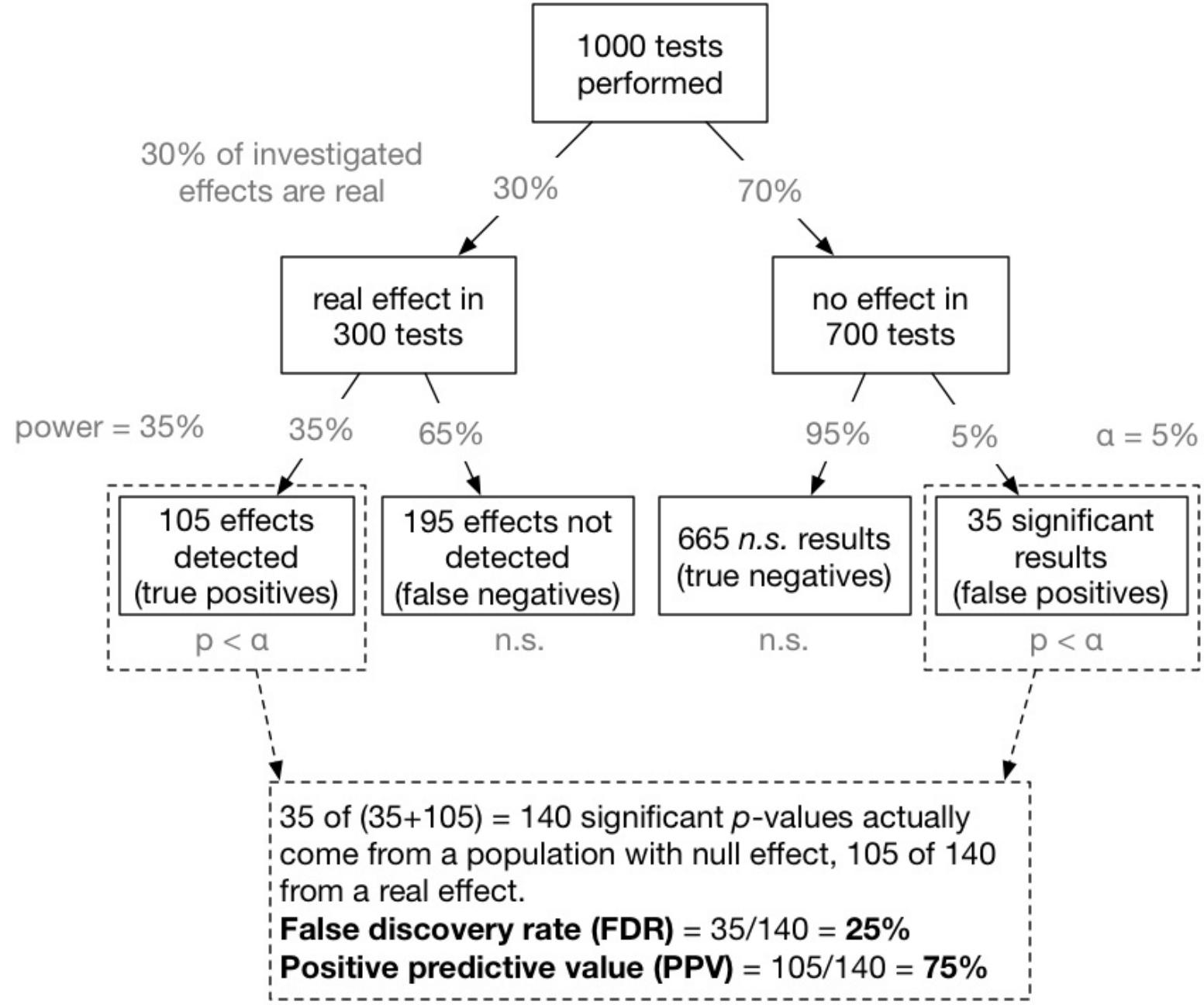
The Neyman model.

Here, we are drawing at random from a box with N tickets. Each ticket represents one unit in the natural-experimental study group. Here, T_i and C_i are the potential outcomes under treatment and control, respectively. If unit i is sampled into treatment, we observe T_i but not C_i ; if unit i is assigned to control, we observe C_i but not T_i . The average of the T_i s in the treatment group estimates the average of all the T_i s in the box, while the average of the C_i s in the control group estimates the average of all the C_i s.

Problems

Random assignment is the “gold standard” for causal inference, but can be misleading under certain circumstances

- Small sample sizes
- Researcher degrees of freedom
- Publication bias
- P-hacking



Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a p -value less than 0.05. Research is not most appropriately represented and summarized by p -values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R + 1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α . Assuming that c relationships are being probed in the field, the expected values of the 2×2 table are given in Table 1. After a research

Copyright © 2005 John P. A. Ioannidis

Study 1: musical contrast and subjective age

In Study 1, we investigated whether listening to a children's song induces an age contrast, making people feel older. In exchange for payment, 30 University of Pennsylvania undergraduates sat at computer terminals, donned headphones, and were randomly assigned to listen to either a control song ("Kalimba," an instrumental song by Mr. Scruff that comes free with the Windows 7 operating system) or a children's song ("Hot Potato," performed by The Wiggles).

After listening to part of the song, participants completed an ostensibly unrelated survey: They answered the question "How old do you feel right now?" by choosing among five options (*very young, young, neither young nor old, old, and very old*). They also reported their father's age, allowing us to control for variation in baseline age across participants.

An analysis of covariance (ANCOVA) revealed the predicted effect: People felt older after listening to "Hot Potato"

(adjusted $M = 2.54$ years) than after listening to the control song (adjusted $M = 2.06$ years), $F(1, 27) = 5.06, p = .033$.

In Study 2, we sought to conceptually replicate and extend Study 1. Having demonstrated that listening to a children's song makes people feel older, Study 2 investigated whether listening to a song about older age makes people *actually* younger.

Study 2: musical contrast and chronological rejuvenation

Using the same method as in Study 1, we asked 20 University of Pennsylvania undergraduates to listen to either "When I'm Sixty-Four" by The Beatles or "Kalimba." Then, in an ostensibly unrelated task, they indicated their birth date (mm/dd/yyyy) and their father's age. We used father's age to control for variation in baseline age across participants.

An ANCOVA revealed the predicted effect: According to their birth dates, people were nearly a year-and-a-half younger after listening to "When I'm Sixty-Four" (adjusted $M = 20.1$ years) rather than to "Kalimba" (adjusted $M = 21.5$ years), $F(1, 17) = 4.92, p = .040$.

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Psychological Science
22(11) 1359–1366
© The Author(s) 2011
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797611417632
<http://pss.sagepub.com>


Joseph P. Simmons¹, Leif D. Nelson², and Uri Simonsohn¹

¹The Wharton School, University of Pennsylvania, and ²Haas School of Business, University of California, Berkeley

Abstract

In this article, we accomplish two things. First, we show that despite empirical psychologists' nominal endorsement of a low rate of false-positive findings ($\leq .05$), flexibility in data collection, analysis, and reporting dramatically increases actual false-positive rates. In many cases, a researcher is more likely to falsely find evidence that an effect exists than to correctly find evidence that it does not. We present computer simulations and a pair of actual experiments that demonstrate how unacceptably easy it is to accumulate (and report) statistically significant evidence for a false hypothesis. Second, we suggest a simple, low-cost, and straightforwardly effective disclosure-based solution to this problem. The solution involves six concrete requirements for authors and four guidelines for reviewers, all of which impose a minimal burden on the publication process.

Table I. Likelihood of Obtaining a False-Positive Result

Researcher degrees of freedom	Significance level		
	$p < .1$	$p < .05$	$p < .01$
Situation A: two dependent variables ($r = .50$)	17.8%	9.5%	2.2%
Situation B: addition of 10 more observations per cell	14.5%	7.7%	1.6%
Situation C: controlling for gender or interaction of gender with treatment	21.6%	11.7%	2.7%
Situation D: dropping (or not dropping) one of three conditions	23.2%	12.6%	2.8%
Combine Situations A and B	26.0%	14.4%	3.3%
Combine Situations A, B, and C	50.9%	30.9%	8.4%
Combine Situations A, B, C, and D	81.5%	60.7%	21.5%

The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time*

Andrew Gelman[†] and Eric Loken[‡]

14 Nov 2013

Abstract

Researcher degrees of freedom can lead to a multiple comparisons problem, even in settings where researchers perform only a single analysis on their data. The problem is there can be a large number of *potential* comparisons when the details of data analysis are highly contingent on data, without the researcher having to perform any conscious procedure of fishing or examining multiple p-values. We discuss in the context of several examples of published papers where data-analysis decisions were theoretically-motivated based on previous literature, but where the details of data selection and analysis were not pre-specified and, as a result, were contingent on data.

Caveats / limitations

Random assignment is the “gold standard” for causal inference, but it has some limitations:

- Randomization often isn’t feasible and/or ethical
- Experiments are costly in terms of time and money
- It’s difficult to create convincing parallel worlds
- Inevitably people deviate from their random assignments

Anyone can flip a coin, but it’s difficult to create convincing parallel worlds

Practical Guide to Controlled Experiments on the Web: Listen to Your Customers not to the HiPPO

Ron Kohavi

Microsoft

One Microsoft Way
Redmond, WA 98052

ronnyk@microsoft.com

Randal M. Henne

Microsoft

One Microsoft Way
Redmond, WA 98052

rhenne@microsoft.com

Dan Sommerfield

Microsoft

One Microsoft Way
Redmond, WA 98052

dans@microsoft.com

5.2 Trust and Execution

5.2.1 Run Continuous A/A Tests

Run A/A tests (see Section 3.1) and validate the following.

1. Are users split according to the planned percentages?
2. Is the data collected matching the system of record?
3. Are the results showing non-significant results 95% of the time?

Continuously run A/A tests in parallel with other experiments.

5.2.2 Automate Ramp-up and Abort

As discussed in Section 3.3, we recommend that experiments ramp-up in the percentages assigned to the Treatment(s). By doing near-real-time analysis, experiments can be auto-aborted if a treatment is statistically significantly underperforming relative to the Control. An auto-abort simply reduces the percentage of users assigned to a treatment to zero. By reducing the risk in exposing many users to egregious errors, the organization can make bold bets and innovate faster. Ramp-up is quite easy to do in online environments, yet hard to do in offline studies. We have seen no mention of these practical ideas in the literature, yet they are extremely useful.

5.2.3 Determine the Minimum Sample Size

Decide on the statistical power, the effect you would like to detect, and estimate the variability of the OEC through an A/A test. Based on this data you can compute the minimum sample size needed for the experiment and hence the running time for your web site. A common mistake is to run experiments that are underpowered. Consider the techniques mentioned in Section 3.2 point 3 to reduce the variability of the OEC.

5.2.4 Assign 50% of Users to Treatment

One common practice among novice experimenters is to run new variants for only a small percentage of users. The logic behind that decision is that in case of an error only few users will see a bad treatment, which is why we recommend Treatment ramp-up. In order to maximize the power of an experiment and minimize the running time, we recommend that 50% of users see each of the variants in an A/B test. Assuming all factors are fixed, a good approximation for the multiplicative increase in running time for an A/B test relative to 50%/50% is $1/(4p(1-p))$ where the treatment receives portion p of the traffic. For example, if an experiment is run at 99%/1%, then it will have to run about 25 times longer than if it ran at 50%/50%.

5.2.5 Beware of Day of Week Effects

Even if you have a lot of users visiting the site, implying that you could run an experiment for only hours or a day, we strongly recommend running experiments for at least a week or two, then continuing by multiples of a week so that day-of-week effects can be analyzed. For many sites the users visiting on the weekend represent different segments, and analyzing them separately may lead to interesting insights. This lesson can be generalized to other time-related events, such as holidays and seasons, and to different geographies: what works in the US may not work well in France, Germany, or Japan.

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer^{a,1}, Jamie E. Guillory^{b,2}, and Jeffrey T. Hancock^{b,c}

^aCore Data Science Team, Facebook, Inc., Menlo Park, CA 94025; and Departments of ^bCommunication and ^cInformation Science, Cornell University, Ithaca, NY 14853

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)

Significance

We show, via a massive ($N = 689,003$) experiment on Facebook, that emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness. We provide experimental evidence that emotional contagion occurs without direct interaction between people (exposure to a friend expressing an emotion is sufficient), and in the complete absence of nonverbal cues.

Editorial Expression of Concern and Correction

PSYCHOLOGICAL AND COGNITIVE SCIENCES

PNAS is publishing an Editorial Expression of Concern regarding the following article: “Experimental evidence of massive-scale emotional contagion through social networks,” by Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock, which appeared in issue 24, June 17, 2014, of *Proc Natl Acad Sci USA* (111:8788–8790; first published June 2, 2014; 10.1073/pnas.1320040111). This paper represents an important and emerging area of social science research that needs to be approached with sensitivity and with vigilance regarding personal privacy issues.

Questions have been raised about the principles of informed consent and opportunity to opt out in connection with the research in this paper. The authors noted in their paper, “[The work] was consistent with Facebook’s Data Use Policy, to which all users agree prior to creating an account on Facebook, constituting informed consent for this research.” When the authors prepared their paper for publication in PNAS, they stated that: “Because this experiment was conducted by Facebook, Inc. for internal purposes, the Cornell University IRB [Institutional Review Board] determined that the project did not fall under Cornell’s Human Research Protection Program.” This statement has since been [confirmed by Cornell University](#).

Obtaining informed consent and allowing participants to opt out are best practices in most instances under the US Department of Health and Human Services Policy for the Protection of Human Research Subjects (the “[Common Rule](#)”). Adherence to the Common Rule is [PNAS policy](#), but as a private company Facebook was under no obligation to conform to the provisions of the Common Rule when it collected the data used by the authors, and the Common Rule does not preclude their use of the data. Based on the information provided by the authors, PNAS editors deemed it appropriate to publish the paper. It is nevertheless a matter of concern that the collection of the data by Facebook may have involved practices that were not fully consistent with the principles of obtaining informed consent and allowing participants to opt out.

Inder M. Verma
Editor-in-Chief

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Correction for “Experimental evidence of massive-scale emotional contagion through social networks,” by Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock, which appeared in issue 24, June 17, 2014, of *Proc Natl Acad Sci USA* (111:8788–8790; first published June 2, 2014; 10.1073/pnas.1320040111).

The authors note that, “At the time of the study, the middle author, Jamie E. Guillory, was a graduate student at Cornell University under the tutelage of senior author Jeffrey T. Hancock, also of Cornell University (Guillory is now a postdoctoral fellow at Center for Tobacco Control Research and Education, University of California, San Francisco, CA 94143).” The author and affiliation lines have been updated to reflect the above changes and a present address footnote has been added. The online version has been corrected.

The corrected author and affiliation lines appear below.

**Adam D. I. Kramer^{a,1}, Jamie E. Guillory^{b,2},
and Jeffrey T. Hancock^{b,c}**

^aCore Data Science Team, Facebook, Inc., Menlo Park, CA 94025; and Departments of ^bCommunication and ^cInformation Science, Cornell University, Ithaca, NY 14853

¹To whom correspondence should be addressed. Email: akramer@fb.com.

²Present address: Center for Tobacco Control Research and Education, University of California, San Francisco, CA 94143.

www.pnas.org/cgi/doi/10.1073/pnas.1412583111

Natural experiments

Natural experiments

Sometimes we get lucky and nature effectively runs experiments for us, e.g.:

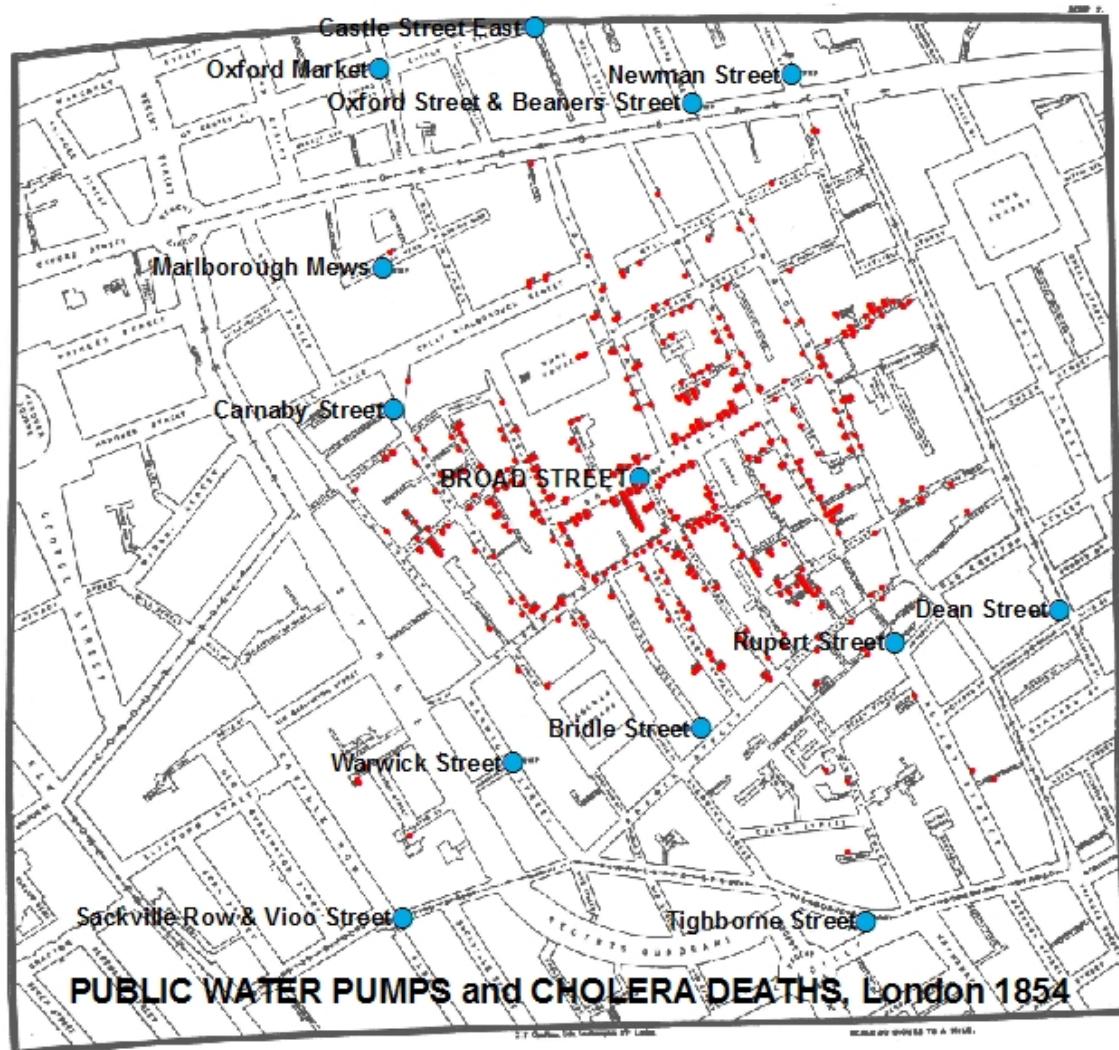
- As-if random: People are randomly exposed to water sources
- Instrumental variables: A lottery influences military service
- Discontinuities: Star ratings get arbitrarily rounded
- Difference in differences: Minimum wage changes in just one state

Natural experiments

Sometimes we get lucky and nature effectively runs experiments for us, e.g.:

- As-if random: People are randomly exposed to water sources
- Instrumental variables: A lottery influences military service
- Discontinuities: Star ratings get arbitrarily rounded
- Difference in differences: Minimum wage changes in just one state

Experiments happen all the time, we just have to notice them



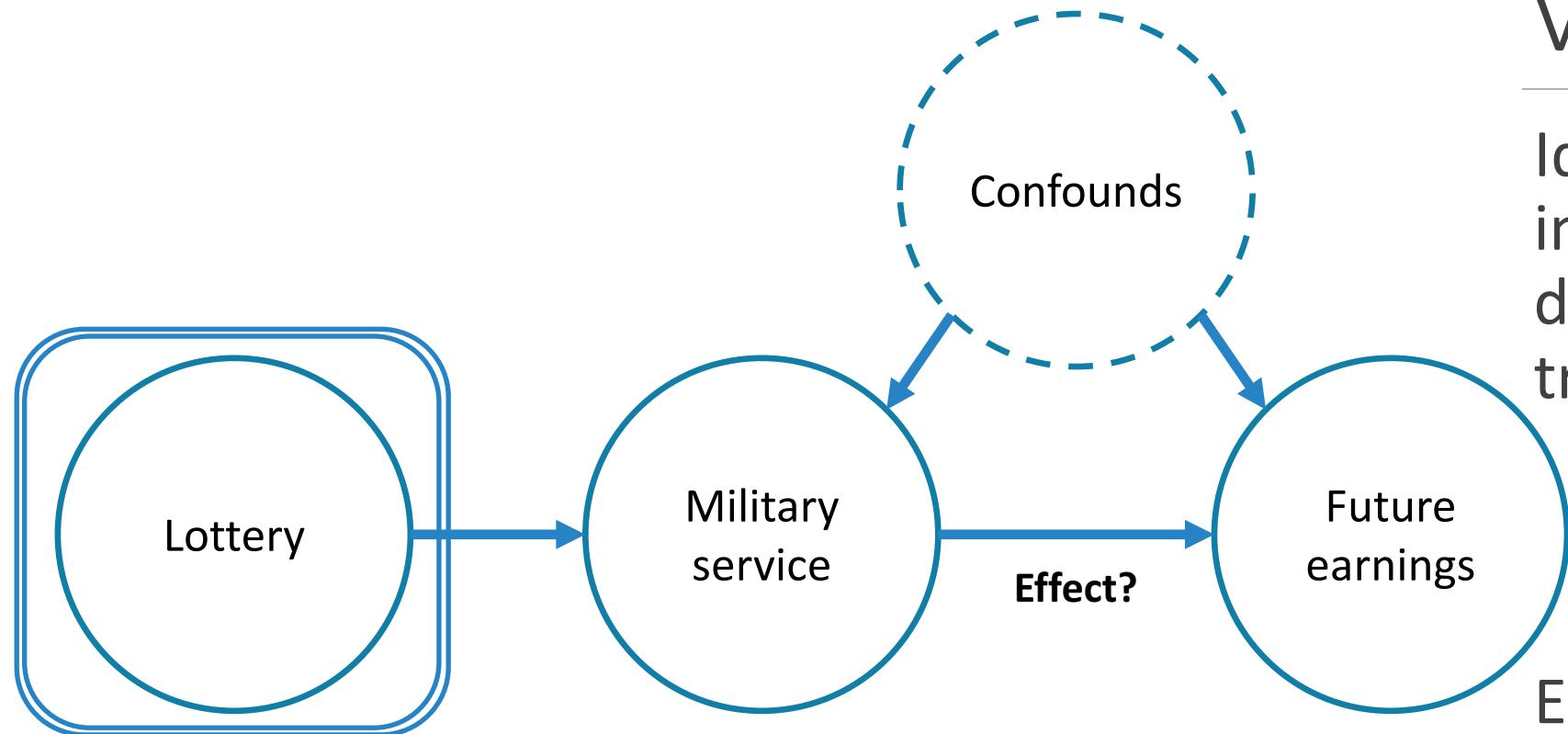
As-if random

Idea: Nature randomly assigns conditions

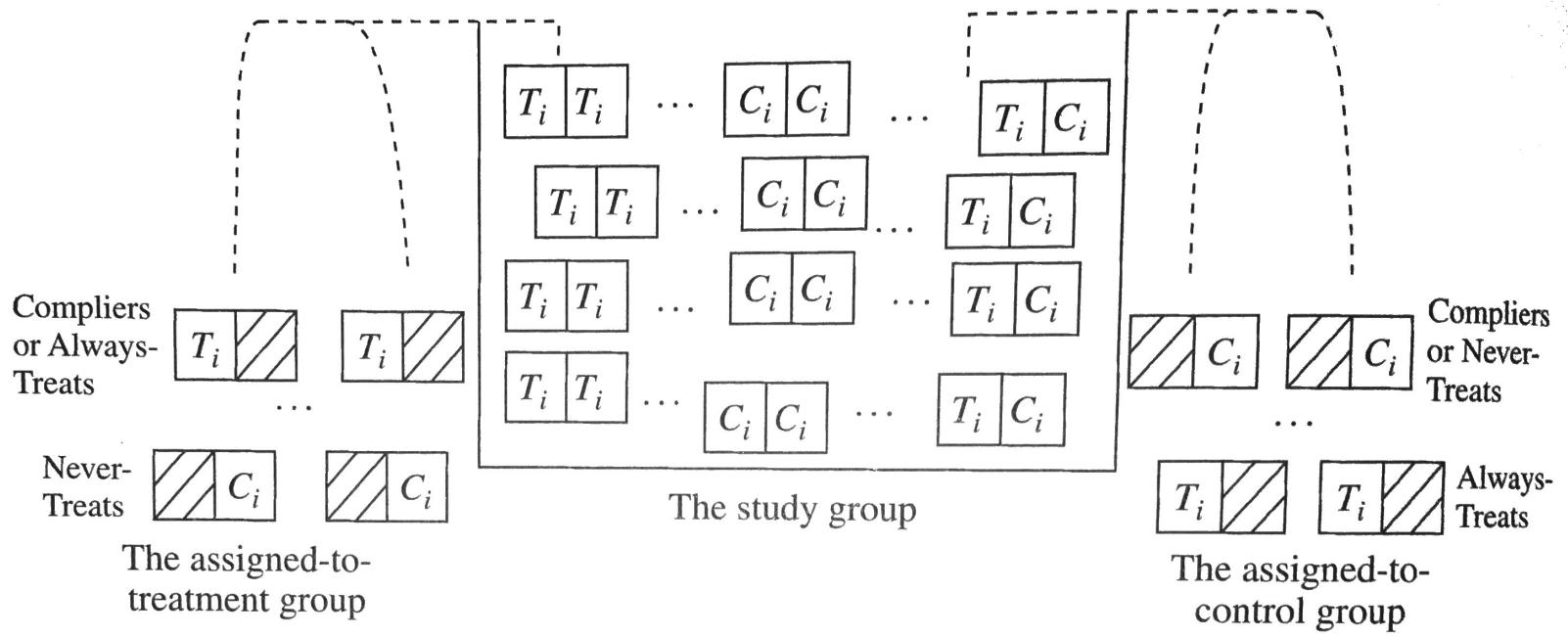
Example: People are randomly exposed to water sources (Snow, 1854)

Instrumental variables

Idea: An instrument independently shifts the distribution of a treatment



Example: A lottery influences military service (Angrist, 1990)



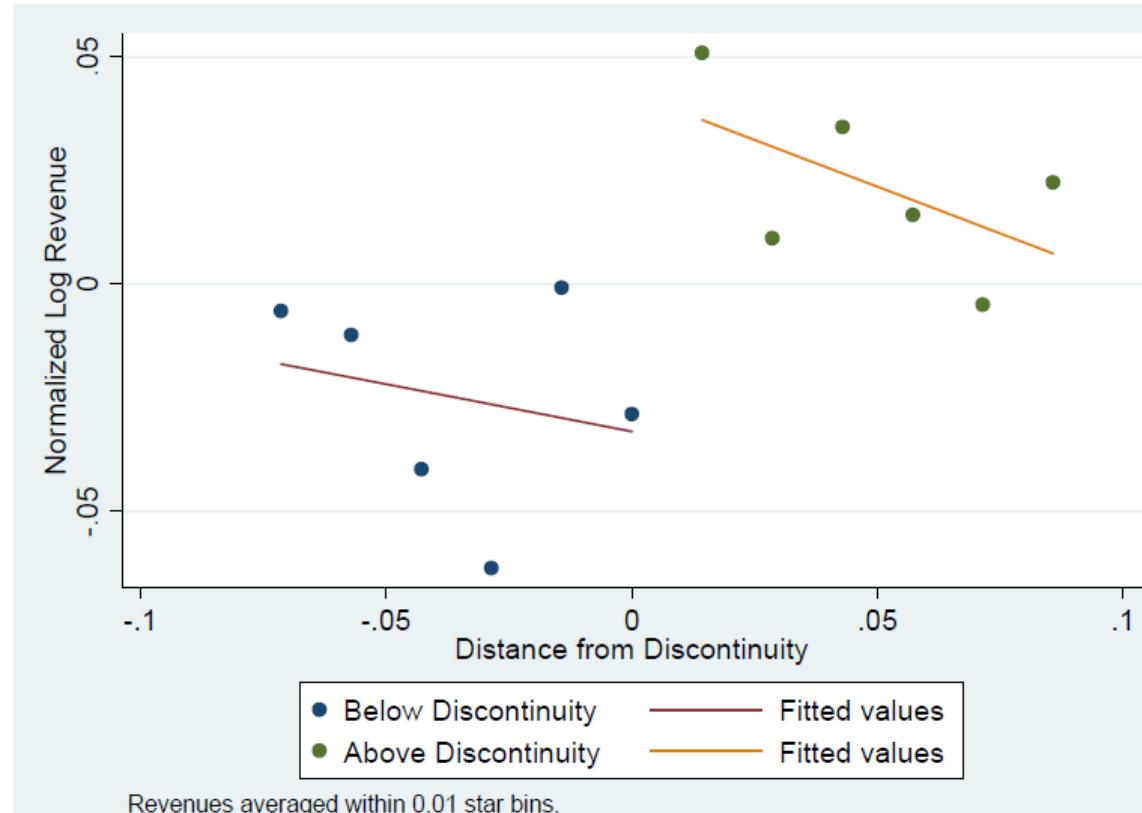
Noncompliance under the Neyman model.

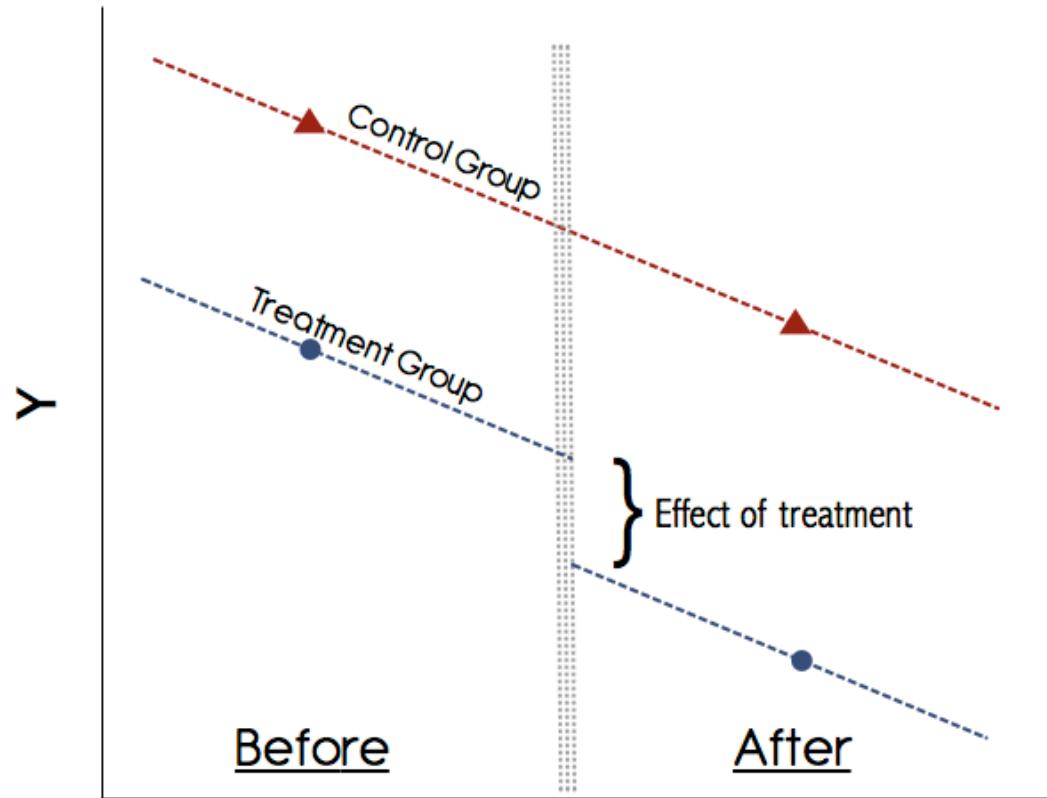
A model of natural-experimental crossover. Each ticket in the box represents one unit in the natural-experimental study group. Every ticket has two fields, one representing potential outcomes under assignment to treatment and the other potential outcomes under assignment to control. Tickets with T_i in both fields are “Always-Treats.” Tickets with C_i in both fields are “Never-Treats.” Tickets with T_i in one field and C_i in the other are “Compliers.” (Defiers are ruled out by assumption.) Here, we draw at random without replacement from a box with N tickets, placing $n < N$ tickets in the assigned-to-treatment group and $m = N - n$ tickets in the assigned-to-control group. The assigned-to-treatment groups and assigned-to-control groups contain a mixture of Always-Treats, Compliers, and Never-Treats; the mixture should be the same in both groups, up to random error, because both groups are random samples of the tickets in the box.

Regression discontinuities

Idea: Things change around an arbitrarily chosen threshold

Example: Star ratings get arbitrarily rounded (Luca, 2011)





Difference in differences

Idea: Compare differences after a sudden change with trends in a control group

Example: Minimum wage changes in just one state (Card & Krueger, 1994)

Natural experiments: Caveats

Natural experiments are great, but:

- Good natural experiments are hard to find
- They rely on many (untestable) assumptions
- The treated population may not be the one of interest

Natural experiments: Caveats

Natural experiments are great, but:

- Good natural experiments are hard to find
- They rely on many (untestable) assumptions
- The treated population may not be the one of interest

Sometimes we can use *additional data + algorithms* to
automatically find natural experiments

Discovering natural experiments

Example: How much traffic to recommender systems *cause*?

(Sharma, Hofman & Watts, 2015)

The screenshot shows the product page for 'Causality: Models, Reasoning and Inference' by Judea Pearl on Amazon. The page includes the book cover, customer reviews (4.5 stars from 18 reviews), and purchase options (Buy New for \$45.87 or Rent for \$32.11). Below the main product information, there is a section titled 'Customers Who Bought This Item Also Bought' which lists several other books related to causal inference and statistics. A red circle highlights this section.

Causality: Models, Reasoning and Inference Hardcover – September 14, 2009
by Judea Pearl (Author)
4.5 stars, 18 customer reviews
ISBN-13: 978-0521895606 | ISBN-10: 052189560X | Edition: 2nd

Buy New
Price: \$45.87

Rent
Price: \$32.11

40 New from \$45.87 | 25 Used from \$38.86

Rent from Amazon Price New from Used from

—	\$30.49	—	—
\$32.11	\$45.87	\$45.87	\$38.86

eTextbook

Hardcover

Written by one of the preeminent researchers in the field, this book provides a comprehensive exposition of modern analysis of causation. It shows how causality has grown from a nebulous concept into a mathematical theory with significant applications in the fields of statistics, artificial intelligence, economics, philosophy, cognitive science, and the health and social sciences. Judea Pearl presents and unifies the probabilistic, manipulative, counterfactual, and structural approaches to causation and devises simple mathematical tools for studying the relationships between causal connections and statistical associations. The book will open the way for including causal analysis in

[Read more](#)

Customers Who Bought This Item Also Bought

Counterfactuals and Causal Inference:... Stephen L. Morgan Paperback \$31.01 Prime	Counterfactuals and Causal Inference:... Stephen L. Morgan Paperback \$37.99 Prime	Probabilistic Reasoning in Intelligent... Judea Pearl Paperback \$52.77 Prime	Probability Theory: The Logic of Science E. T. Jaynes Hardcover \$91.21 Prime	Mostly Harmless Econometrics: An Empiricist's Companion Joshua D. Angrist, Andrew Gelman Paperback \$24.54 Prime	Bayesian Data Analysis, Third Edition (Chapman & Hall/CRC Texts in Statistical Science) Andrew Gelman, John Hill, Jennifer Hill Hardcover \$60.52 Prime	Causal Inference for Statistics, Social, and Clinical Research: An Introduction Guido W. Imbens, Donald W. Rubin Hardcover \$56.15 Prime	An Introduction to Causal Inference Judea Pearl Paperback \$9.99 Prime
---	---	--	--	---	--	---	---

Observational estimates

Naively, up to 30% of pageviews come through recommendations

March 2009

A photograph showing a woman from behind, looking at a computer monitor. The monitor displays a large red rectangular object, possibly a speaker or a piece of equipment. In the bottom left corner of the image, there is a small portrait of a man with glasses and a nameplate below it.

**Jeffrey Grau,
Senior Analyst
jgrau@emarketer.com**

**Personalized Product
Recommendations:
*Predicting Shoppers' Needs***

Executive Summary: Personalized recommendations help shoppers discover products of interest when there is a large selection. Only a minority of online retailers use recommendation systems—but many of those that do report sales increases of 10% or more, as well as higher customer engagement and loyalty.

Observational estimates

Naively, up to 30% of pageviews come through recommendations

- Burton Snowboard, a sports retailer, reported that personalized product recommendations have driven nearly 25% of total sales since it began offering them in 2008. Prior to this, Burton's customer recommendations consisted of items from its list of top-selling products.

Observational estimates

Naively, up to 30% of pageviews come through recommendations

- Burton Snowboard, a sports retailer, reported that personalized product recommendations have driven nearly 25% of total sales since it began offering them in 2008. Prior to this, Burton's customer recommendations consisted of items from its list of top-selling products.

But this is almost surely an overestimate of the effect

Typical browsing

WITH RECOMMENDATIONS

Typical browsing: Search

The screenshot shows the Amazon search results for "winter hat". The top navigation bar includes the Amazon logo, a search bar with the query "Men's Accessories winter hat", and links for "Your Account", "Prime", "Wish List", and a shopping cart. Below the search bar, there are category links like "Amazon Fashion", "Women", "Men", "Girls", "Boys", "Baby", "Luggage", "Sales And Deals", and "Your Lists". A banner at the top says "Mother's Day Savings". The main content area displays 1-60 of 12,135 results for "Clothing, Shoes & Jewelry : Men : Accessories : \"winter hat\"". The results are arranged in a grid of 4 columns and 3 rows. Each item has a thumbnail image, the product name, price range, Prime status, and a star rating. A red circle highlights the fourth item in the first row: "Wigwam Mills Inc F4707 Worsted Wool Watch Cap". The bottom of the page features a "Refine by Brand" section with checkboxes for various brands.

Typical browsing: Focal product

amazon Try Prime Your Amazon.com Today's Deals Gift Cards Sell Help

Shop by Department Search Clothing, Shoes & Jewelry winter hat

Amazon Fashion Women Men Girls Boys Baby Luggage Sales And Deals Your

Mother's Day Savings

FREE RETURNS

◀ Back to search results for "winter hat"

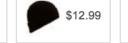





Wigwam
Wigwam Mills Inc F4707 Worsted Wool Watch Cap
★★★★★ 600 customer reviews

List Price: \$17.99
Price: **\$12.99** & FREE Shipping on orders over \$35. FREE Returns. Details
You Save: \$5.00 (28%)

Size: One Size [Size Chart](#) | Fit: As expected (77%)

Color: Grey
 \$12.99  \$9.88  \$9.59  \$12.99  \$9.25  \$12.99

In Stock.
Sold by Cambridge Select and Fulfilled by Amazon. Gift-wrap available.

Want it tomorrow, April 30? Order within 9 hrs 5 mins and choose One-Day Shipping at checkout. Details

- Worsted Wool
- Machine Wash Warm Water; Tumble Dry Low Heat
- Made from thick & heavy Worsted Wool
- Durable, quality beanie style watchcap
- Keeps your head and ears protected even in negative temperatures
- Made in the USA!
- One Size Fits Most

Frequently Bought Together

 Rothco \$5.61 - \$8.95

 Carhartt \$7.96 - \$17.85

 Duofold \$8.94 - \$29.70

 People Socks \$24.95

Share    

Qty: 1

Yes, I want FREE Two-Day Shipping with Amazon Prime

Turn on 1-Click ordering for this browser

Ship to:
CA 94301

Add to Wish List

Other Sellers on Amazon

\$14.99
FREE Shipping on orders over \$35.00 & FREE Returns. Details
Sold by: CuffCrazy LLC

3 new from \$12.99

Typical browsing: Recommended product

Amazon.com - Today's Deals - Gift Cards - Sell - Help

Shop by Department - Search: Sports & Outdoors

Mother's Day Savings

Sports & Outdoors - Sports & Fitness - Outdoor Recreation - Sports Fan Shop - Sports Deals - Outdoor Deals

Clothing, Shoes & Jewelry > Men

Rothco

Black Military Wool Glove Liners

★★★★★ 268 customer reviews

List Price: \$8.99
Price: **\$7.59** & FREE Shipping on orders over \$35. [Details](#)
You Save: **\$1.40** (16%)

Size: Medium

Size Chart

Color: Black

In Stock.

Sold by [Built For Survival](#) and Fulfilled by Amazon. Gift-wrap available.

Want it tomorrow, April 30? Order within 5 hrs 50 mins and choose One-Day Shipping at checkout. [Details](#)

- Great To Protect Your Fingers
- Made In The U.S.A
- Made From 70% Wool & 30% Nylon Material
- Comfortable and Warm
- Can Be Used As A Liner or Alone As A Glove

Frequently Bought Together

Wigwam \$9.25 - \$14.99	Rothco \$10.64	Army Universe \$9.99	Rothco \$1.99 - \$37.22

Roll over image to zoom in

Share: Email, Facebook, Twitter, Pinterest

Qty: 1

Yes, I want FREE Two-Day Shipping with [Amazon Prime](#)

Add to Cart

Turn on 1-Click ordering for this browser

Ship to: CA 94301

Add to Wish List

Other Sellers on Amazon

\$5.99 + Free Shipping Sold by: OutdoorBunker
\$5.99 + Free Shipping Sold by: Go Commando Gear
\$6.30 + Free Shipping Sold by: Army Universe

11 new from \$5.61

Counterfactual browsing

WITHOUT RECOMMENDATIONS

Counterfactual browsing: Search

amazon Try Prime Your Amazon.com Today's Deals Gift Cards Sell Help

Shop by Department Search Men's Accessories winter hat

Mother's Day Savings

Hello, Sign in Your Account Try Prime Wish List Cart

FREE RETURNS

1-60 of 12,135 results for Clothing, Shoes & Jewelry : Men : Accessories : "winter hat"

Show results for Any Department Clothing, Shoes & Jewelry Men Accessories Hats & Caps Earmuffs Gloves & Mittens Scarves Belts Sunglasses & Eyewear Accessories Bow Ties & Cummerbunds Neckties Cuff Links, Shirt Studs & Tie Clips Clothing Watches Shoes Jewelry Shops Uniforms, Work & Safety Surf, Skate & Street Big & Tall Refine by Brand

City Hunter W430 Premium Denim Trapper Hat (4 Colors)
\$8.99 - \$17.99 ✓Prime
★★★★★ 158

Best Winter Hats 40 Gram Thinsulate Insulated Cuffed Winter Hat (One Size)
\$8.99 - \$11.99 ✓Prime
★★★★★ 97

LOCOMO Tentacle Octopus Cthulhu Knit Beanie Hat Cap Wind Ski Mask FFFH135DBLU
\$5.21 - \$21.73 \$29.99
★★★★★ 36

Wigwam Mills Inc F4707 Worsted Wool Watch Cap
\$7.96 - \$17.85 ✓Prime
★★★★★ 2,097

Carhartt Men's Acrylic Watch Hat
\$11.99 \$24.99
★★★★★ 30

ODEMA Unisex Nylon Russian Style Winter Ear Flap Hat
\$11.99 \$24.99
★★★★★ 30

Columbia

Carhartt

NEFF

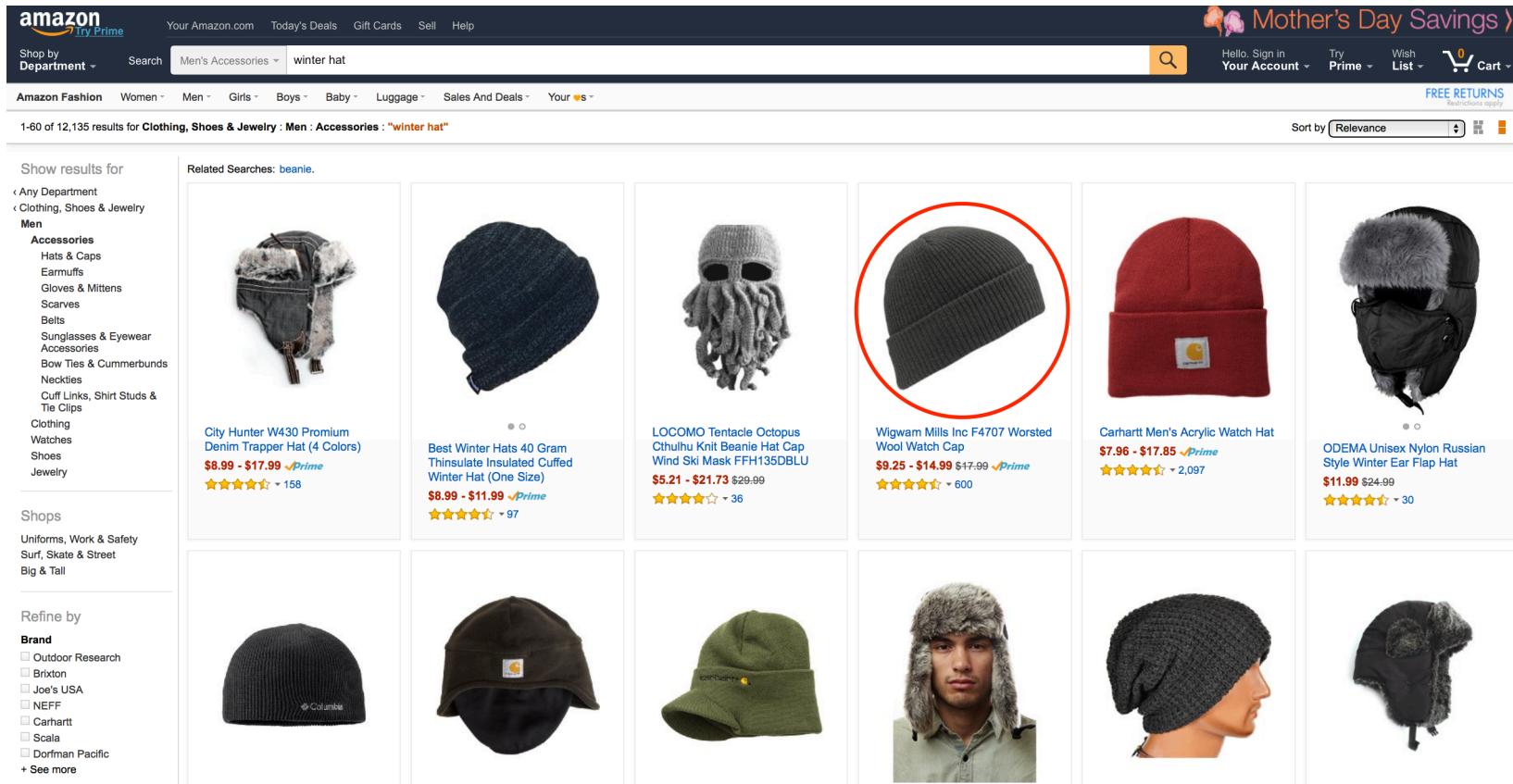
Brixton

Joe's USA

Scala

Dorfman Pacific

+ See more



Counterfactual browsing: Focal product

amazon Try Prime Your Amazon.com Today's Deals Gift Cards Sell Help

Shop by Department Search Clothing, Shoes & Jewelry winter hat

Amazon Fashion Women Men Girls Boys Baby Luggage Sales And Deals Your

Mother's Day Savings

FREE RETURNS

◀ Back to search results for "winter hat"

Wigwam
Wigwam Mills Inc F4707 Worsted Wool Watch Cap

★★★★★ 600 customer reviews

List Price: \$17.99
Price: \$12.99 & FREE Shipping on orders over \$35. FREE Returns. Details
You Save: \$5.00 (28%)

Size: One Size | Size Chart | Fit: As expected (77%)

Color: Grey

\$12.99 \$9.88 \$9.59 \$12.99 \$9.25 \$12.99

In Stock.
Sold by Cambridge Select and Fulfilled by Amazon. Gift-wrap available.

Want it tomorrow, April 30? Order within 9 hrs 5 mins and choose One-Day Shipping at checkout. Details

- Worsted Wool
- Machine Wash Warm Water; Tumble Dry Low Heat
- Made from thick & heavy Worsted Wool
- Durable, quality beanie style watchcap
- Keeps your head and ears protected even in negative temperatures
- Made in the USA!
- One Size Fits Most

Roll over image to zoom in

Share Email Facebook Twitter Pinterest

Qty: 1 Add to Cart

Yes, I want FREE Two-Day Shipping with Amazon Prime

Turn on 1-Click ordering for this browser

Ship to: CA 94301 Add to Wish List

Other Sellers on Amazon

\$14.99 Add to Cart
FREE Shipping on orders over \$35.00 & FREE Returns. Details
Sold by: CuffCrazy LLC

3 new from \$12.99



Counterfactual browsing: Search again

amazon try Prime

Your Amazon.com Today's Deals Gift Cards Sell Help

Shop by Department Search Men's Accessories wool men's gloves

Mother's Day is May 10 ›

Hello, Sign in Your Account Try Prime Wish List Cart FREE RETURNS Refund policies apply

Sort by Relevance

1-60 of 1,680 results for Clothing, Shoes & Jewelry : Men : "wool men's gloves"

Show results for

Any Department
Clothing, Shoes & Jewelry
Men
Accessories
Clothing
Shoes

Shops
Uniforms, Work & Safety
Surf, Skate & Street
Big & Tall

Refine by

Brand

- People Socks
- Wigwam
- Thirty 48
- Terramar
- Minus33 Merino Wool
- Carhartt
- New Era

+ See more

Eligible for Free Shipping

Free Shipping by Amazon

Avg. Customer Review

- ★★★★★ & Up
- ★★★★☆ & Up
- ★★★☆☆ & Up
- ★★☆☆☆ & Up

International Shipping

- AmazonGlobal Eligible

Price

Showing most relevant results. See all results for [wool men's gloves](#).



[Black Military Wool Glove Liners](#)
\$5.61 - \$8.95 \$8.99
★★★★☆ 268



[Fox River Four Layer Glove](#)
\$22.09 - \$44.00 ✓Prime
★★★★☆ 125



[Fox River Men's Mid Weight Ragg Glove](#)
\$6.50 - \$14.81 ✓Prime
★★★★☆ 30



[Rothco G.I. Wool Glove Liners](#)
\$4.75 - \$8.13 \$9.99
★★★★☆ 136



[Outdoor Research Men's Biosensor Liners](#)
\$19.95 - \$149.15 ✓Prime
★★★★☆ 100



[Outdoor Research Men's Flurry Gloves](#)
\$22.80 - \$75.95 ✓Prime
★★★★☆ 60



[SilverTip Merino Wool gloves](#)



[Fox River Ragg and Leather](#)

Counterfactual browsing

Amazon.com - Today's Deals - Gift Cards - Sell - Help

Shop by Department - Search: Sports & Outdoors

Mother's Day Savings

Sports & Outdoors - Sports & Fitness - Outdoor Recreation - Sports Fan Shop - Sports Deals - Outdoor Deals

Clothing, Shoes & Jewelry > Men

Rothco

Black Military Wool Glove Liners

★★★★★ 268 customer reviews

List Price: \$8.99
Price: **\$7.59** & FREE Shipping on orders over \$35. [Details](#)
You Save: **\$1.40** (16%)

Size: Medium

Size Chart

Color: Black

In Stock.

Sold by [Built For Survival](#) and Fulfilled by Amazon. Gift-wrap available.

Want it tomorrow, April 30? Order within 5 hrs 50 mins and choose One-Day Shipping at checkout. [Details](#)

- Great To Protect Your Fingers
- Made In The U.S.A
- Made From 70% Wool & 30% Nylon Material
- Comfortable and Warm
- Can Be Used As A Liner or Alone As A Glove

Roll over image to zoom in

Share: Email, Facebook, Twitter, Pinterest

Qty: 1

Yes, I want FREE Two-Day Shipping with [Amazon Prime](#)

Add to Cart

Turn on 1-Click ordering for this browser

Ship to: CA 94301

Add to Wish List

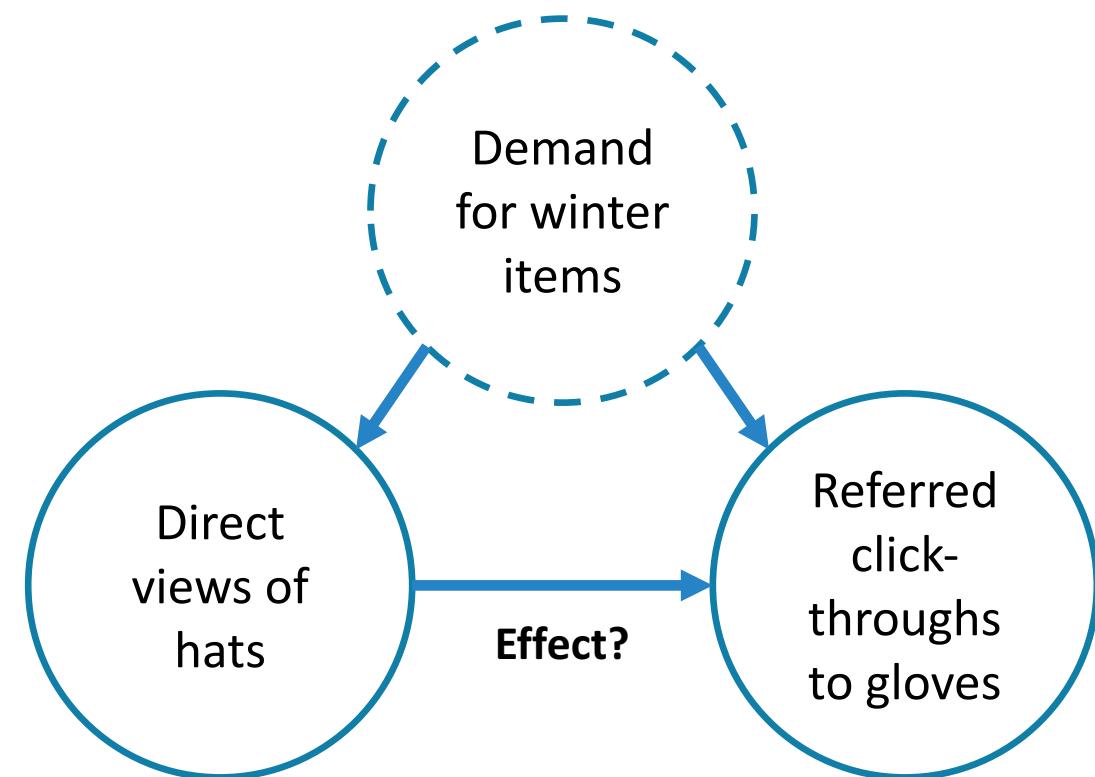
Other Sellers on Amazon

\$5.99	Add to Cart
+ Free Shipping	
Sold by: OutdoorBunker	
\$5.99	Add to Cart
+ Free Shipping	
Sold by: Go Commando Gear	
\$6.30	Add to Cart
+ Free Shipping	
Sold by: Army Universe	
11 new from \$5.61	

Confound: Correlated demand

Some views *would have happened anyway* due to correlated demand

We call these *convenience clicks*



Ideal experiment

Ideally we would run an A/B test where randomly selected people see recommendations

This screenshot shows the Amazon product page for the Wigwam Mills Inc F4707 Worsted Wool Watch Cap. The product image is a dark grey ribbed beanie. The page includes a 'Mother's Day Savings' banner, a search bar, and navigation links. The product title is 'Wigwam Mills Inc F4707 Worsted Wool Watch Cap'. It has a 4.5-star rating from 600 reviews. The price is \$12.99, with a note about free shipping on orders over \$35. The product is in stock and made in the USA. A 'Frequently Bought Together' section shows related items like gloves, carhartt pants, duofold socks, and people socks.

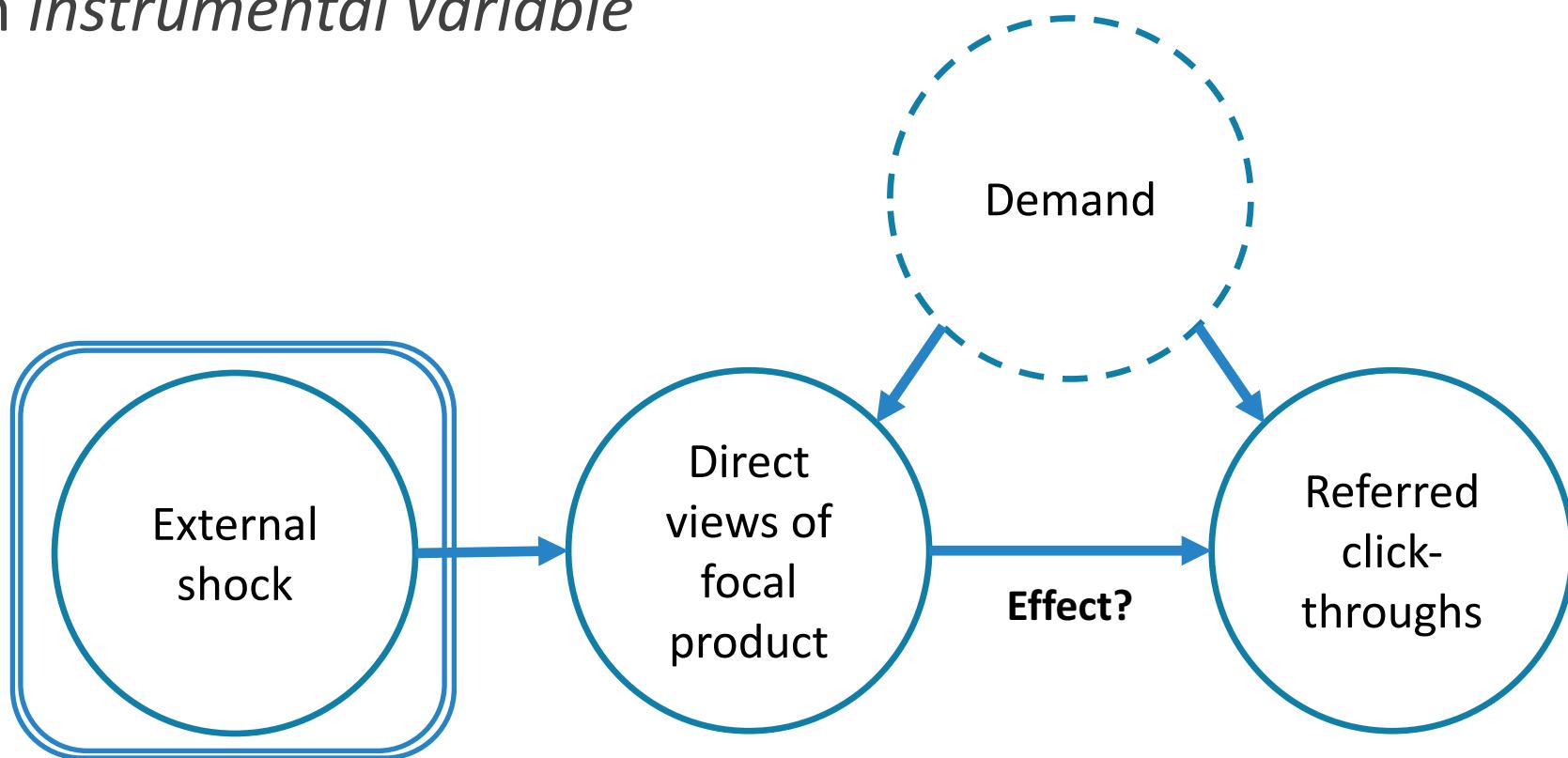
World 1

This screenshot shows the same Amazon product page for the Wigwam Mills Inc F4707 Worsted Wool Watch Cap, but with different visual elements. The product image is a dark grey ribbed beanie. The page includes a 'Mother's Day Savings' banner, a search bar, and navigation links. The product title is 'Wigwam Mills Inc F4707 Worsted Wool Watch Cap'. It has a 4.5-star rating from 600 reviews. The price is \$12.99, with a note about free shipping on orders over \$35. The product is in stock and made in the USA. A 'Frequently Bought Together' section shows related items like gloves, carhartt pants, duofold socks, and people socks.

World 2

Natural experiment

Instead, we can exploit *sudden shocks in traffic* to focal products as an *instrumental variable*



Your Amazon.com Today's Deals Gift Cards Sell Help

Shop by Department Search All causal inference

Books Advanced Search New Releases Best Sellers The New York Times® Best Sellers Children's Books Textbooks Textbook Rentals Sell Us Your Books Best Books of the Month

Books > Science & Math > Mathematics > Applied > Probability & Statistics

Causality: Models, Reasoning and Inference Hardcover – September 14, 2009
by Judea Pearl (Author)
★★★★★ 18 customer reviews
ISBN-13: 978-0521895606 | ISBN-10: 052189560X | Edition: 2nd

Buy New Price: \$45.87
40 New from \$45.87 | 25 Used from \$38.86

Rent Price: \$32.11
Rent from — \$30.49
Amazon Price \$32.11 \$45.87

eTextbook
Hardcover

Written by one of the preeminent researchers in the field, this book provides a comprehensive exposition of modern analysis of causation. It shows how causality has grown from a nebulous concept into a mathematical theory with significant applications in the fields of statistics, artificial intelligence, economics, philosophy, cognitive science, and the health and social sciences. Judea Pearl presents and unifies the probabilistic, manipulative, counterfactual, and structural approaches to causation and devises simple mathematical tools for studying the relationships between causal connections and statistical associations. The book will open the way for including causal analysis in

[+ Read more](#)

CAUSALITY
SECOND EDITION
MODELS, REASONING, AND INFERENCE
JUDEA PEARL

See this image

Customers Who Bought This Item Also Bought

Counterfactuals and Causal Inference:... Stephen L. Morgan ★★★★★ 7 Paperback \$31.01	Counterfactuals and Causal Inference:... Stephen L. Morgan ★★★★★ 4 Paperback \$37.99	Probabilistic Reasoning in Intelligent Systems Judea Pearl ★★★★★ 11 Paperback \$52.77	Probability Theory: The Logic of Science E. T. Jaynes ★★★★★ 42 Hardcover \$91.21	Mostly Harmless Econometrics: An Empiricist's Companion Joshua D. Angrist and E. T. Jaynes ★★★★★ 42 Paperback \$24.54	Bayesian Data Analysis, Third Edition (Chapman & Hall/CRC Texts in Statistical Science) Andrew Gelman and Jennifer Hill ★★★★★ 25 Hardcover \$60.52	Causal Inference in Statistics, Society, and the Sciences Guido W. Imbens and Donald W. Rubin ★★★★★ 1 Hardcover \$56.15

Usual approach

Think hard for a source of random variation that only directly affects focal product (e.g., author wins award)

amazon Try Prime

Your Amazon.com Today's Deals Gift Cards Sell Help

Shop by Department Search All causal inference

Books Advanced Search New Releases Best Sellers The New York Times® Best Sellers Children's Books Textbooks Textbook Rentals Sell Us Your Books Best Books of the Month

Books > Science & Math > Mathematics > Applied > Probability & Statistics

Causality: Models, Reasoning and Inference Hardcover – September 14, 2009
by Judea Pearl (Author)
★★★★★ 18 customer reviews
ISBN-13: 978-0521895606 | ISBN-10: 052189560X | Edition: 2nd

Buy New Price: \$45.87
40 New from \$45.87 | 25 Used from \$38.86

Rent Price: \$32.11
Rent from — \$30.49
Amazon Price \$32.11 \$45.87

eTextbook Hardcover

Written by one of the preeminent researchers in the field, this book provides a comprehensive exposition of modern analysis of causation. It shows how causality has grown from a nebulous concept into a mathematical theory with significant applications in the fields of statistics, artificial intelligence, economics, philosophy, cognitive science, and the health and social sciences. Judea Pearl presents and unifies the probabilistic, manipulative, counterfactual, and structural approaches to causation and devises simple mathematical tools for studying the relationships between causal connections and statistical associations. The book will open the way for including causal analysis in

[+ Read more](#)

See this image

Customers Who Bought This Item Also Bought

- Counterfactuals and Causal Inference:...
Stephen L. Morgan
★★★★★ 7 Paperback \$31.01 ✓Prime
- Counterfactuals and Causal Inference:...
Stephen L. Morgan
★★★★★ 4 Paperback \$37.99 ✓Prime
- Probabilistic Reasoning in Intelligent...
Judea Pearl
★★★★★ 11 Paperback \$52.77 ✓Prime
- Probability Theory: The Logic of Science
E. T. Jaynes
★★★★★ 31 Hardcover \$91.21 ✓Prime
- Mostly Harmless Econometrics
Joshua D. Angrist
★★★★★ 42 Paperback \$24.54 ✓Prime
- Bayesian Data Analysis, Third Edition (Chapman...
Andrew Gelman
★★★★★ 25 Hardcover \$60.52 ✓Prime
- Causal Inference: Discrete and Continuous Models
Guido W. Imbens
★★★★★ 1 Hardcover \$56.15 ✓Prime

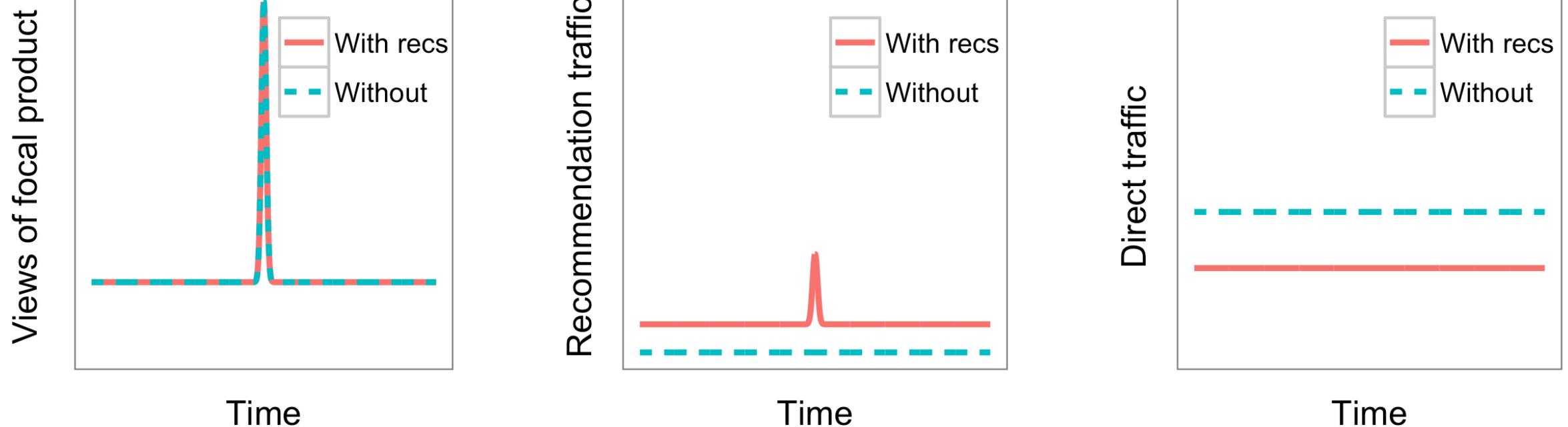
Usual approach

Think hard for a source of random variation that only directly affects focal product (e.g., author wins award)

Problem: Impossible to rule out side effects

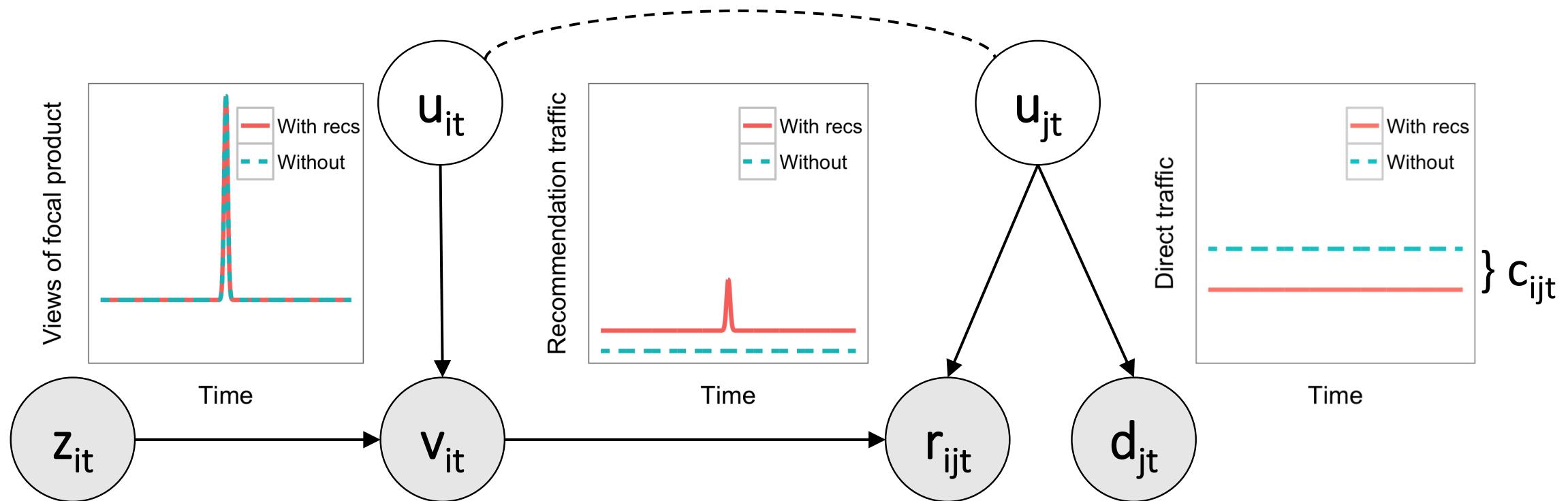
New approach: Automatically discovering natural experiments

Look for products that receive shocks in direct traffic, while their recommendations do not



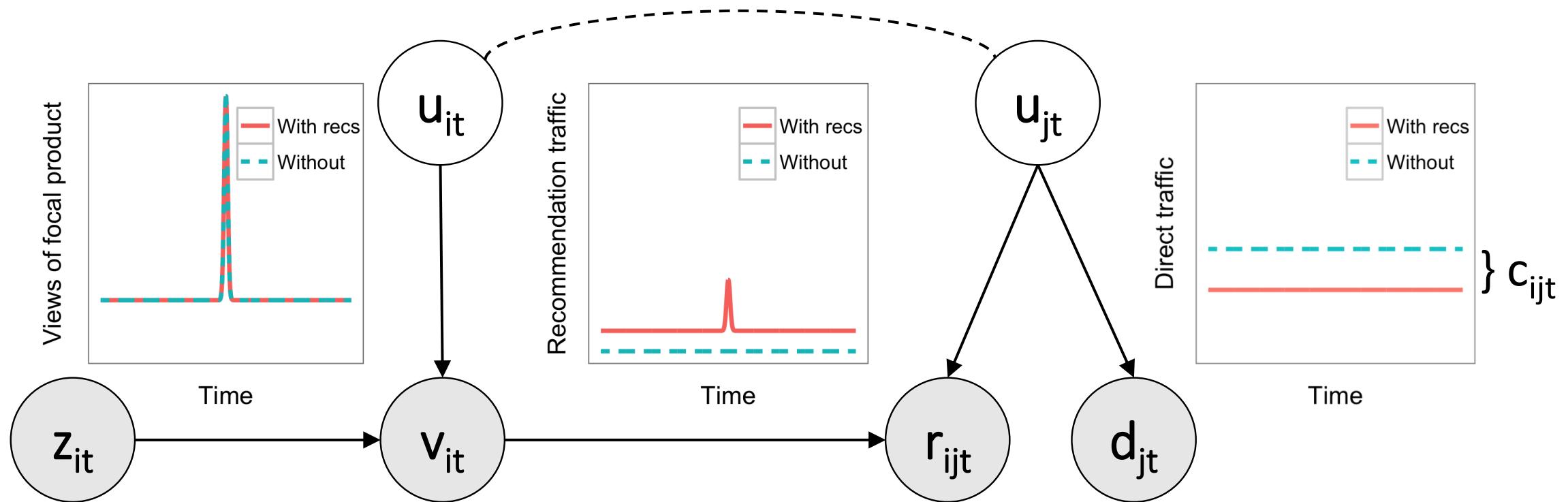
New approach: Automatically discovering natural experiments

Look for products that receive shocks in direct traffic, while their recommendations do not



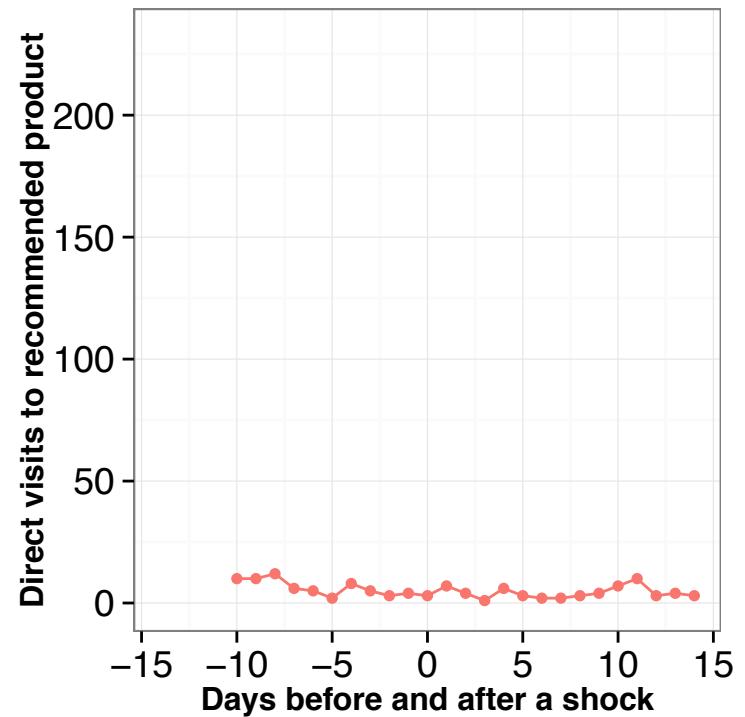
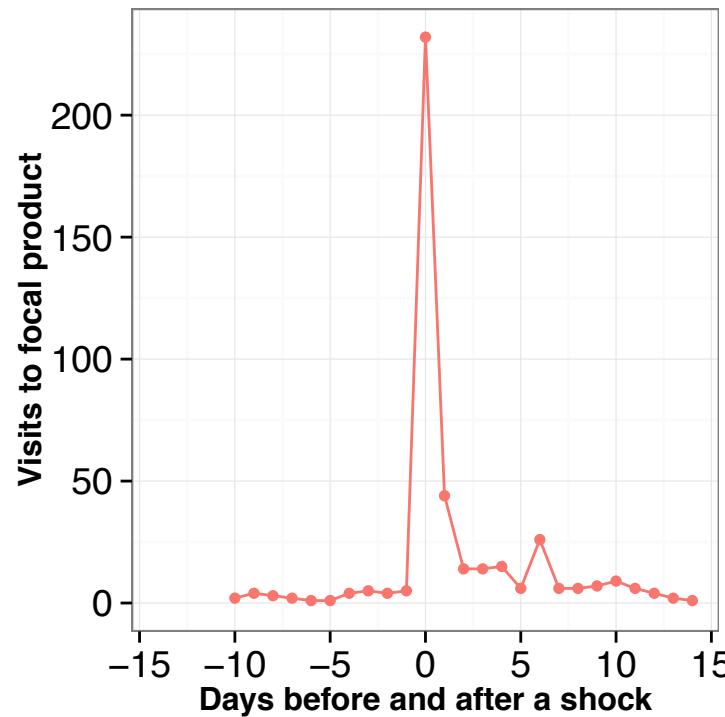
New approach: Automatically discovering natural experiments

Look for products that receive shocks in direct traffic, while their recommendations do not



New approach: Automatically discovering natural experiments

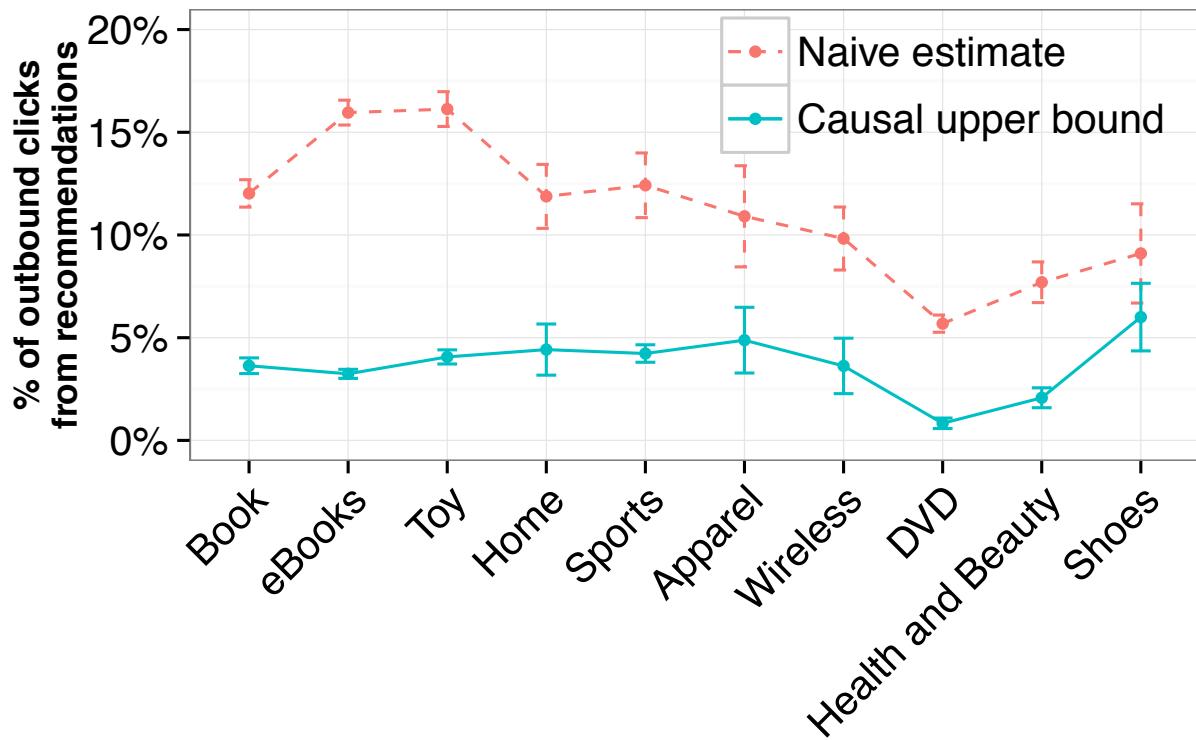
Applying this method to 23 million pageviews in the Bing Toolbar logs, we find over 4,000 such natural experiments



New approach: Automatically discovering natural experiments

Causal click-through rate is just the marginal gain in clicks during the shock

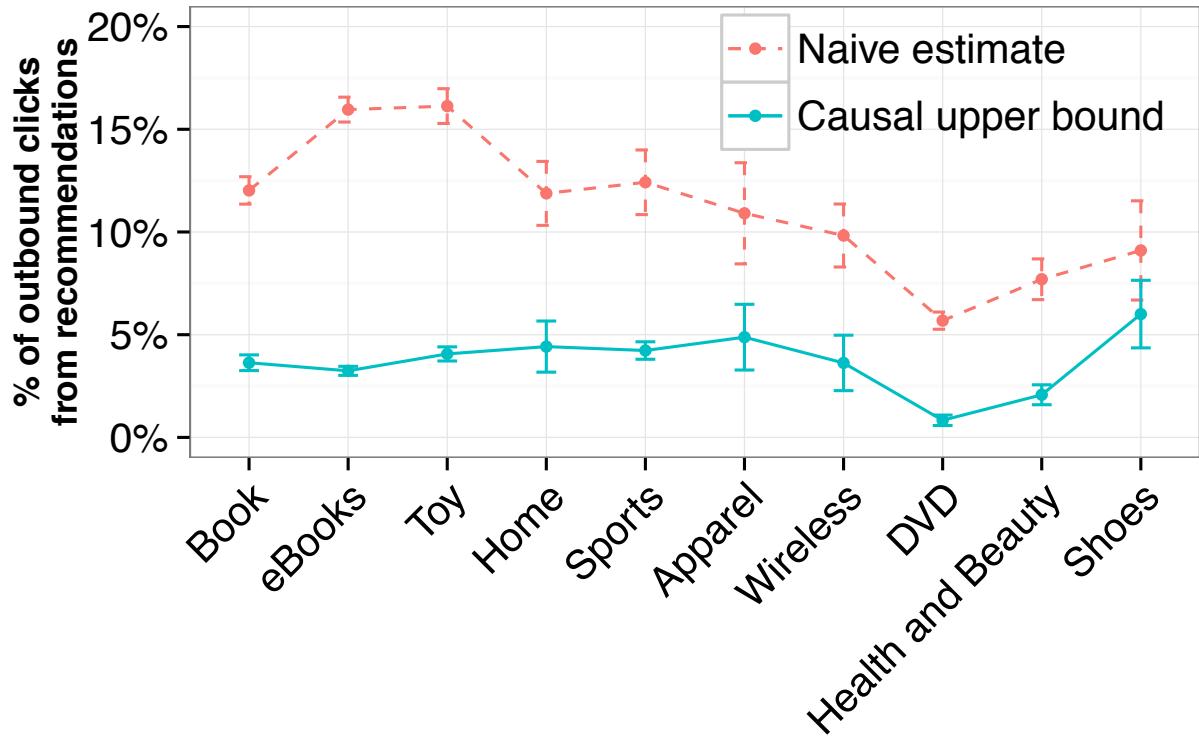
Causal effect = Change in recommendation clicks / Size of shock



Results

Causal click-through rate is just the marginal gain in clicks during the shock

Causal effect = Change in rec clicks /
Size of shock



Results

Although recommendation click-throughs account for a large fraction of traffic, *at least 75% of this activity would likely occur in the absence of recommendations**

*With lots of caveats

Closing thoughts

Large-scale *observational data* is useful for building *predictive models* of a *static world*

Closing thoughts

But without appropriate *random variation*, it's hard to
predict what happens when you change something in the
world

Closing thoughts

Randomized experiments are like *custom-made datasets* to answer a specific question

Closing thoughts

Additional data + algorithms can help us discover and analyze these examples in the wild