abcNEWS.com | WHY? | abcNEWS ON DEMAND

powered by [hp]

**Good Morning America**   **World News Tonight**   **20/20**   **Primetime**   **Nightline**   **UpClose**   **WNN**

February 2, 2003

**SCI/TECH**

HOMEPAGE
NEWS SUMMARY
US
INTERNATIONAL
MONEYScope
WEATHER
LOCAL NEWS
ENTERTAINMENT
ESPN SPORTS
SCI / TECH
POLITICS
HEALTH
TRAVEL

**FEATURED SERVICES**
SHOPPING
DOWNLOADS
WIRELESS

**INTERACT**
VIDEO & AUDIO
BOARDS
CHAT
NEWS ALERTS
CONTACT ABC

abcNEWS ON DEMAND
**VIEW IT**
abcNEWSstore.com

abcNEWS.com  O r i g i n a l   R e p o r t

Worried about sharing information online? There may be a mathematical solution to helping people safeguard their privacy.

**SEARCH**

# Private Clicks

**Mathematical Balance Between Public Statistics, Personal Information**

*Commentary*
*By John Allen Paulos*
*Special to ABCNEWS.com*

**Feb. 2 — You want to buy a rug or a calendar online. Or maybe you want to subscribe to some Internet publication. You fill in your address and credit card number, but the web site, hungry for information about its visitors, goes on to ask you your age, your income, your hobbies, or any of a number of other questions that are of questionable relevance to its business.**
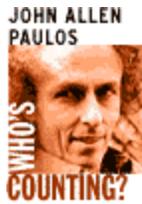
[icon] **Print This Page**
[icon] **Email This Page**
[icon] **See Most Sent**

ALSO ON abcNEWS.com

● **Why Is the Shuttle Debris Dangerous?**

● **How Will Disaster Affect NASA's Future?**

● **History of Deadly Space Accidents**

MORE ON THIS STORY

> **RELATED STORIES**

● **Who's Counting?: Privacy and Terrorists**

JOHN ALLEN PAULOS
WHO'S COUNTING?

What many, if not most, people do in such situations is lie. You claim to be 97, to make $4 million a year, and to have a keen interest in ice fishing in Honolulu.

The companies making these requests sometimes have their reasons. They want to target their ads, improve services, respond to customers' desires, and for all this it helps if they have a rough idea of who's buying their products.

Recently Rakesh Agrawal and Ramakrishnan Srikant, two IBM researchers in California, have developed a simple program that might make customer duplicity less appealing. Based on the realization that companies often want data that is roughly accurate in the aggregate but not necessarily personally revealing, their program partially reconciles companies' desire for information with individuals' desire for privacy.

## The Program

Here's how it handles a nosy question: You answer it honestly, the program generates a random number that is either added to or subtracted from this answer, and only this last number is sent to the company. Using straightforward statistical techniques, the company can still recover approximate averages and correlations from the numbers submitted, and this is often sufficient for their purposes.

For example, someone indicates that she is 46 years old and the program adds or subtracts some random number between zero and 20 to this age so that the number the company receives might be 60 or 33 or any number between 26 and 66. Likewise with

**SCI / TECH**

● **Search f... Disaster**

● **Mathema... Maintaining...**

● **In Wake ... Science May...**

● **Shuttle D...**

income figures. Someone says he makes $120,000 a year and some number between zero and $50,000 is added to or subtracted from this number before it is submitted to the company.

There is a trade-off. The numbers are more valuable and accurate if the random fudge factor added or subtracted to the correct answer is small. The smaller the number, however, the less protection afforded the individual. Another problem is that the customer has to trust the company enough not to record the individual random number used to mask the exact answer.

There are variations possible involving the distribution of the numbers added or subtracted. They can be adjusted to accommodate customers' residual lying in somewhat the same way that people, knowing that others generally want to appear younger and richer than they are, interpret statements about age and income.

The mathematics comes in when trying to reconstruct the true averages and correlations from all the more or less false individual numbers. The program uses Bayes' theorem, a powerful result in probability (that was implicit in January's column on terrorists and privacy), to help in this reconstruction.

### Have You X-ed? Another Example

The idea of obtaining demographic and other information without compromising personal privacy has been around for a long time. For a different sort of illustration, let's assume we have a large group of people and we want to discover what percentage of them have done something, say X, that they'll probably be embarrassed to admit. Assume also that there is a legitimate reason, say medical or otherwise, for our wanting to know what percentage of people have X-ed. What can we do?

Again we use a randomizing device and ask each person in the group to flip a coin and keep the result secret. If the coin lands heads, the person is instructed to answer honestly: Has he or she ever X-ed — Yes or No? If the coin lands tails, the person is instructed to simply answer Yes. Thus a Yes response could mean one of two things, one quite innocuous (the coin landing tails), the other potentially embarrassing (X-ing). Since no one can know what the Yes means, presumably people will be honest.

For illustration, let's say that 560 of 1000 responses are Yes. What does this indicate about the percentage of people who have X-ed? Approximately 500 of the 1000 people will answer Yes simply because their coin landed tails. We'll ignore them and focus only on the 500 people whose coin landed heads and who therefore replied to the question honestly. Of these 500 approximately 60 answered Yes. Thus 12 percent (60/500) is the estimate for the percentage of people who have X-ed. There are many refinements of this method that can be used to learn more detail, such as how many times people have X-ed.

Until such techniques become widespread, lying is a reasonable strategy when confronted with overly probing questions. Time for my Hawaiian ice fishing lesson. ■

*Professor of mathematics at Temple University and adjunct professor of journalism at Columbia University, John Allen Paulos is the author of several best-selling books, including* Innumeracy*, and the forthcoming* A Mathematician Plays the Market *, which will be published in the spring. His* Who's Counting? *column on ABCNEWS.com appears the first weekend of every month.*

PRINT THIS PAGE    |    SEND THIS TO A FRIEND    |    VIEW MOST EMAILED