

Valuating Privacy

In several experimental auctions, participants put a dollar value on private information before revealing it to a group. An analysis of results show that a trait's desirability in relation to the group played a key role in the amount people demanded to publicize private information.



BERNARDO A. HUBERMAN, EYTAN ADAR, AND LESLIE R. FINE
Hewlett-Packard Labs

Because people can easily obtain, aggregate, and disperse personal data electronically, privacy is a central concern in the information age. This concern is clear in relation to financial data and genetic information, both of which can lead to identity abuse and discrimination. However, other relatively harmless information can also be abused, including a person's gender, salary, age, marital status, or shopping preferences. What's unclear is whether it's the fear of such abuse that actually causes people's stated hesitance to reveal their data.

Several researchers have used survey techniques to reveal correlations between individual self-disclosures and demographic data.^{1,2} The Jourard Self-Disclosure Questionnaire, for example, shows to whom people will disclose information, but fails to capture the data's specific value. More recent work further clarifies the privacy trade-offs that individuals will make to gain access to specific services, pointing out the disparities between stated privacy attitudes and actions.³⁻⁵ However, researchers have yet to quantify how much people value their secrets and to what extent that valuation depends on context. (By *context*, we mean the make-up of the group viewing the data.) Our hypothesis—and the motivation for our study—is that people reveal information when they feel that they're somewhat typical or positively atypical compared to the target group.

To test this hypothesis, we conducted experiments that elicit the value people place on their private data. Specifically, our initial question was the following: Is a trait's desirability or undesirability the primary factor in dictating how a person values a piece of information? We found, with great significance (more than 95 percent statistical

confidence) that a linear relationship exists between an individual's belief about a trait and the value he or she places on it. That is, the less desirable the trait, the greater the price a person demands for releasing the information. Furthermore, we found that small deviations in a socially positive direction are associated with a lower asking price.

Experimental design

By treating private information as a real good,⁶ we designed our economic experiment to determine information's value by offering to purchase it from subjects and reveal it to a group. This, in effect, elicited the individual's "privacy calculus."

We told subjects that they would participate in a reverse second-price auction for personal data: the individual demanding the least for the information was paid the second-lowest demanded price. In exchange for this compensation, and after verification, the individual had to reveal that piece of information to the other auction participants. The auction's financially competitive nature, coupled with the fact that all participants had to anonymously submit their private data along with their demanded price, let us extract the value that each individual placed on disclosing the private information.

Data collection

We conducted 10 separate test sessions with 127 individuals (59 percent male, 41 percent female) recruited through local colleges and company mailing lists. Five of the sessions were mixed gender, three were female only, and two were male only. In all sessions, we conducted a weight auction;

we also held the age auction in seven of the 10 sessions, with 88 participants total (57 percent male). In a post-experiment questionnaire, we presented all participants with hypothetical bidding scenarios on financial data and also asked them questions about their attitudes toward privacy, their own weight, and their beliefs about other session participants, including how many they knew and how well they knew them.

We paid participants a nominal fee (US\$25) for attending, along with anything they earned in the auction. In all auctions, we limited prices to US\$100 or less; “infinity” indicated that US\$100 would not be enough for the person to reveal the information to others. We gave all subjects a randomly assigned identifier and kept no records that linked individuals to their number. After we fully explained the experiment to the subjects, we asked them to sign a consent form. Participants were free to leave or not participate at any time. Each auction form and the survey indicated the participant’s ID number.

For the weight auction, we collected gender, height, price, and weight information. For age auctions, we collected only age, gender, and price. We declared the bid with the lowest price the winner; when there were ties, we randomly selected the winner from among the tying bids. We validated weight using a scale and age using participants’ driver’s licenses. To enforce truthful revelation, we required subjects to be within five pounds of the weight they listed on their bid forms.

Data analysis

We recorded all “infinity” bids to a randomly selected price between US\$100 and US\$2,000, using the log of the price to prevent large variations. We normalized weight according to the Body Mass Index (BMI; calculated as weight/height², in kilograms and centimeters, respectively). To determine statistical significance, we used the Kruskal-Wallis Anova test, and Tukey’s Honestly Significantly Different test for pairwise comparisons of binned data.

Results

Weight and age are examples of private information that people value, that can be verified instantly, and that have no financial or identify-theft repercussions. Of the two, our weight auction, which included 127 participants, displayed the strongest effects.

Weight and age auction results

Figure 1 shows the relationship between weight (binned by percentile) and the log of the prices participants demanded to make that information public. We used a Kruskal-Wallis Anova test to reveal statistical significance ($p = 0.019$) with a distinct visual trend in the average price as a function of BMI. Those individuals weighing slightly below average—an “ideal” weight by cultural

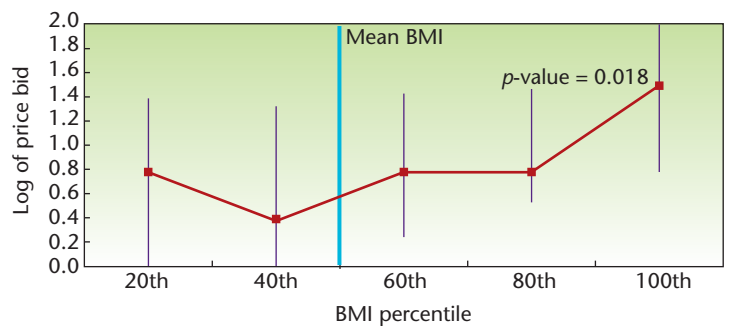


Figure 1. Kruskal-Wallis Anova analysis of the log of price bids in bins based on Body Mass Index (BMI) percentiles.

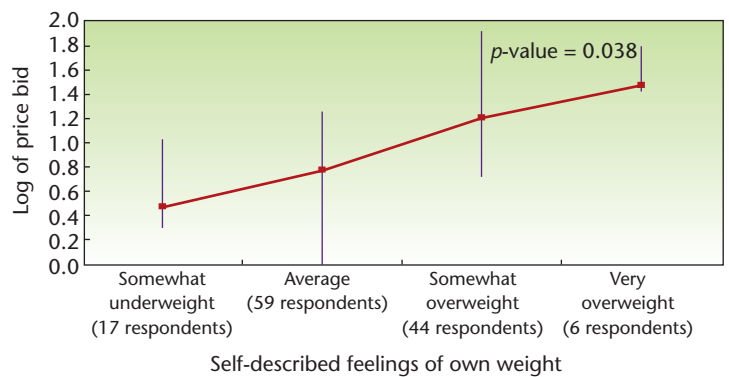


Figure 2. Kruskal-Wallis Anova analysis of the log of price bids in bins based on participant’s perception of their weight.

standards—required little compensation to publicize this fact. In contrast, those who weighed more and might therefore fear embarrassment or stigmatization,⁷ demanded relatively high compensation. Interestingly, although anyone can visually infer a characteristic such as weight, it’s still considered private. Such behavior is potentially linked to our internal (and potentially false) beliefs on how we’re perceived by groups, as well as how we perceive ourselves.^{8,9}

To test the impact of self-perception factors, part of our post-experiment questionnaire asked each subject whether they believed themselves to be “very under,” “somewhat under,” “average,” “somewhat over,” or “very over” in relation to the average weight of the other subjects. Binned by these categories, the results are even more striking than actual weight. Once again, those who perceived themselves to be very underweight indicated that they would reveal weight information for a small amount of money. As perception of weight relative to average increased, so too did the price demanded. As Figure 2 shows, the slightly higher price demanded by the

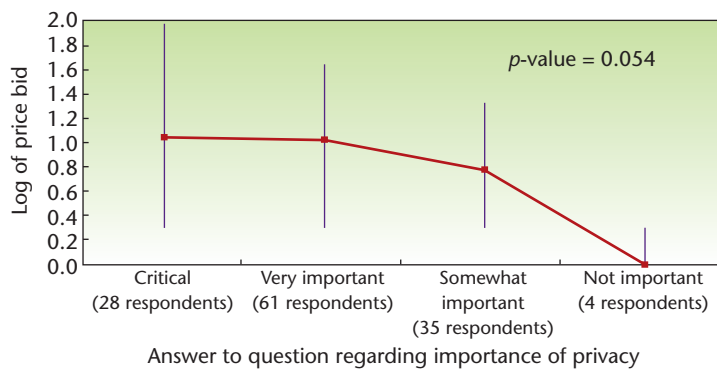


Figure 3. Kruskal-Wallis Anova analysis of the log of price bids, in bins based on participants' privacy attitudes.

lowest weight group in Figure 1 disappears when binning by perception. This suggests that while certain subjects had a low weight in reality, they did not perceive themselves as such and priced their information accordingly.

While weaker than the weight auction trends, our age auction results showed similar tendencies. We gathered age data during seven of the 10 experiments, representing 88 participants with an age range of 23 to 62 years and an average age of 40. We analyzed the log of price bids in bins based on the subject's age and found a slight increase with age ($p = .17$). However, it's notable that for the two extreme bins, we found a significant log price difference (0.665 or \$3.62, versus 1.28 or \$18.05, with $p = .0297$). The implication is that very young subjects are more willing to reveal their age than older ones, and that the large population segments in between are similar in their privacy demands. Unlike the weight auctions, the smaller demand differences of different age group segments might also indicate that age information is less sensitive than weight. This interpretation is supported by the difference in average demand price for the two auctions (\$57.56 for age versus \$74.06 for weight).

Hypothetical auctions

In addition to age and weight, we used hypothetical auctions to study participant price demands for information about salary, spousal salary, credit rating, and savings. As part of the post-experiment survey, we asked all participants to imagine an auction for this data and to indicate how much they would demand.

The percentage of individuals demanding more than US\$100 was 48 percent for salary (77 respondents), 36 percent for spousal salary (52 respondents), 24 percent for credit rating (78 respondents), and 38 percent for savings (77 respondents). All demands are relatively high in comparison to weight and age auctions, where only 5.5 percent and 3.5 percent, respectively, demanded more than \$100. This additional cost might relate in part to social

taboos against openly discussing controversial information such as salary, and indicates that it's critical to set correct auction limits in such cases. Furthermore, in our case, many subjects knew each other in a professional context and might therefore have evaluated the potential future impacts of revealing financial information. Also, just as the BMI normalization was a better metric than raw weight, factors such as occupation, years on the job, and so on might help explain noisy trends in the pricing data.

General attitudes on privacy

We sought to quantify how general privacy attitudes impacted the price participants set for revealing private information. In the post-auction survey, we posed the question, "How important to you is your personal privacy information?" Response options were "critical," "very important," "somewhat important," and "unimportant." Figure 3 shows the weight prices binned by these categories. Although not insignificant ($p = .056$), general privacy attitudes are clearly not as strong as other factors.

Our survey also attempted to discern how many other auction participants each subject knew. In the weight auction, individuals who were in the top 50th percentile in terms of demanded price knew an average of 36 percent of other participants. The bottom 50th percentile knew only 23 percent ($p = .05$), which suggests that individuals are less reluctant to reveal information to an anonymous audience (the "phenomenon of the stranger"⁹). Unfortunately, most subjects knew about the same number of people, so we saw no effect when binning based on the percent of people known versus price. Thus, this effect, given our population, can at best be considered weak.

We found very slight behavioral variations between men and women. In the weight auction, for example, men demanded an average log price of 0.847 (\$6.03), whereas women demanded 1.13 on average (\$12.49, $p = 0.15$). Six of the seven individuals demanding "infinity" were women. Examining the trends in price demanded as a function of perceived weight, both curves display a marked upward trend, but the male trend appears better defined ($p = 0.0037$ for the male trend, $p = 0.2$ for the female). For women, this result might be partially due to response distribution (for example, only one subject considered herself underweight). Female subjects believing themselves to be "average" displayed a broad price variation; comparing only the "somewhat over" to the "somewhat under" is significant at $p = 0.099$.

In addition to measuring how a trait's desirability impacts the value of private information, our results help explain an apparent paradox: despite stated privacy concerns, individuals frequently give away or sell myriad per-

sonal data. To inform debates on privacy issues ranging from financial information¹⁰ to genetic and medical data¹¹ to surveillance,¹² we must carefully consider how individuals choose to reveal their private information. Our results, which highlight this decision's strongly contextual nature, also suggest ways to increase people's comfort levels in revealing private data.

Our experimental methodology is a novel and useful way to elicit true values for private information. Beyond providing a general understanding of privacy, the method can also help survey designers decide the payment levels most likely to elicit responses from a representative population. □

Acknowledgments

We thank Sara Dubowsky, Mette Huberman, and Lada Adamic for their useful comments and suggestions.

References

1. P. Cozby, "Self-Disclosure: A Literature Review," *Psych. Bulletin*, vol. 79, no. 2, 1973, pp. 73–91.
2. S.M. Jourard, *Self-Disclosure: An Experimental Analysis of the Transparent Self*, Wiley-Interscience, 1971.
3. A. Acquisiti and J. Grossklags, "Losses, Gains, and Hyperbolic Discounting: An Experimental Approach to Information Security Attitudes and Behavior," *Proc. 2nd Int'l Workshop Economics and Info. Security*, 2003; www.cpppe.umd.edu/rhsmith3/agenda.htm.
4. I.H. Hann et al., "Overcoming Online Information Privacy Concerns: A Comparison of Privacy Policies, Convenience, and Promotions," working paper, Marshall School of Business, Univ. of Southern California, Sept. 2003; www-rcf.usc.edu/~hann/publications_files/overcoming%20online%20information%20privacy%20concerns.pdf.
5. J. Kleinberg, C.H. Papadimitriou, and P. Raghavan, "On the Value of Private Information," *Proc. 8th Conf. Theoretical Aspects of Rationality and Knowledge (TARK-2001)*, J. van Benthem, ed., Morgan Kaufmann, 2001, pp. 249–257.
6. A.F. Westin, *Privacy and Freedom*, Atheneum, 1970.
7. E. Goffman, *Stigma: Notes on the Management of Spoiled Identity*, Prentice-Hall, 1963.
8. E. Goffman, *The Presentation of Self in Everyday Life*, Doubleday, 1959.
9. G. Simmel, *The Sociology of Georg Simmel*, K.H. Wolf, ed., The Free Press, 1967.
10. S. Singleton, "California's Financial Privacy Debate: Privacy Policy for an Open Economy," open letter, Competitive Enterprise Inst., May 2002; www.cei.org/gencon/027,03004.cfm.
11. F.S. Collins and J.D. Watson, "Genetic Discrimination: Time to Act," *Science*, vol. 302, 2003, p. 745.
12. B. Schneier, "Protecting Privacy and Liberty," *Nature*, vol. 413, 25 Oct. 2001, p. 773.

Bernardo Huberman is a senior Hewlett-Packard fellow and director of HP's Information Dynamics Laboratory. A physicist by training, his work over the past 10 years has focused on understanding distributed processes and designing mechanisms for information aggregation, preserving privacy, and market-based distributed resource allocation systems. Contact him at bernardo.huberman@hp.com.

Eytan Adar is a PhD student in computer science at the University of Washington, and previously worked for Xerox PARC and HP Laboratories. His research interests are in social networks, Web mining, privacy, and graph visualization. Adar has a BS and an ME from the Massachusetts Institute of Technology. Contact him at eadar@cs.washington.edu.

Leslie Fine is a scientist in the Information Dynamics Lab at HP Labs, where she works as an applied game theorist and mechanism designer with an emphasis on experimental economics. Her recent work focuses on nonmarket mechanisms for information aggregation, the nature of privacy, and market-based distributed resource allocation systems. Contact her at leslie.fine@hp.com.



Stay on Track

IEEE Internet Computing reports emerging tools, technologies, and applications implemented through the Internet to support a worldwide computing environment.

In 2005, we'll look at

- Information Discovery: Needles and Haystacks
- Internet Media
- Social Networks: Interactions, Interfaces, and Infrastructures
- Security for P2P and Ad Hoc Networks

... and more!

IEEE
Internet Computing

www.computer.org/internet/