

APRIL 21, 2008, 3:03 PM

Can You Trust Someone Like You?

By *JOHN TIERNEY*

If Groucho Marx wouldn't belong to any club that would let him be a member, would you believe a survey that would let you be part of the sample?

That was the question I posed last week after describing a survey of Lab readers that is analyzed in the current issue of the *Journal of Consumer Research*.

I asked how reliable an online survey of volunteers could be, and some readers wanted no part of it. Ryan Reich pronounced it "useless scientifically" because there's no way to be sure who the respondents were. Gary Ender objected to my reporting data from "self-selected samples."

Here's a reply to that criticism from two of the behavioral economists who published the study, Scott Rick of the University of Pennsylvania and George Loewenstein of Carnegie Mellon University:

It is reasonable to question the generalizability of our findings, given that the majority of our respondents were a self-selected group of TierneyLab readers. Respondents and non-respondents, no doubt, differ on a variety of dimensions, such as how much free time they have, their ability or motivation to introspect, and their desire to learn more about themselves.

But the way in which we recruited respondents affects the generalizability of some findings more than others. Claims about the distribution of tightwads and spendthrifts in the general population are clearly the most vulnerable to the selection critique, and we note in our paper that this distribution depends heavily on the sample from which respondents are drawn. While we found that tightwads outnumbered spendthrifts by a 3-to-2 ratio among TierneyLab respondents, spendthrifts outnumbered tightwads 2-to-1 when we surveyed female shoppers at a Pittsburgh mall. Even among female shoppers, however, a population in which one would expect spendthrifts to be vastly over-represented, there is a sizeable minority of tightwads.

When we look at the relationships between variables, such as between being a tightwad and spending, however, the selection critique is less serious. Our studies used college students, respondents recruited from TierneyLab, and respondents recruited from other media outlets, and we generally observed similar patterns between variables in all three

groups. For example, we found that tightwads from all three groups are more likely to be male, to save more money, and were the most sensitive to factors that influence the pain of paying (whether or not a \$5 fee is framed as “small,” for example). Different groups differ in their proportion of tightwads, but the tightwads in different groups tend to behave in similar, characteristic, ways.

Finally, it is worth noting that almost any single paper in the social science literature is vulnerable to generalizability critiques, whether they use Internet-based surveys like ours or not. Surveys in which a randomly selected, nationally representative sample is invited to participate face the same critique if less than 100% of invited respondents participate. And most laboratory studies are based on a non-representative subset of college students, themselves unlikely to be representative of the general population.

Many classic psychology experiments used only male participants (Stanley Milgram’s study of obedience in 1963 and Leon Festinger and James Carlsmith’s study of cognitive dissonance in 1959, for example). Contemporary fMRI studies only use participants who are not claustrophobic, who are able to stay up late at night (scanner time is often cheapest at night), and, often, who are right-handed. And the extent to which many findings based on American samples generalize to other cultures is unclear since there is little cross-cultural research in the field of judgment and decision-making (though Joseph Henrich and collaborators, among others, have produced some notable exceptions).

As a result, generalizability is typically not attainable in one study or one paper. Over time, however, investigating the same topic with different methods and different samples can provide converging (and generalizable) evidence.

I find the researchers’ arguments persuasive, and I see online surveys as a great new tool for social science because they can gather so much information. Sure, the individual pieces information aren’t as precise as traditional surveys, but when aggregated they can be valuable.

One analogy: The anonymous commenters at Amazon.com aren’t reliable individually, and I realize that some of them are hyping books and other products sold by themselves and their friends. But when there are huge numbers of comments, I find the aggregate ratings to be one of the most useful sources of consumer information. And I’d bet the respondents who respond to online surveys like the one on spending habits are more trustworthy than the Amazon raters, if only because they have no financial incentive to lie.