The growing challenge of survey fraud and how to address it

As a research agency working across multiple countries, vertically, and categories (Consumer and Enterprise), one of the most important things we do is ensure we have legitimate and engaged participants in the fieldwork we conduct every day.

The challenge of ensuring sample integrity and in turn, data quality, is not new – ever since we all started using online sample, we have placed a barrier between ourselves as researchers and our target respondents. While this major technological enhancement enabled our industry to grow, reduce cost per interview and increase fieldwork efficiencies, it also increased the risk of erroneous responses, "low quality" (i.e., disengaged) respondents and fraudulent participants.

"Old" methods used to be enough to identify quality issues

As an agency, we continually focus on things within our control, such as improving survey formats, guarding survey length, pilot testing fieldwork to capture any data collection issues, and running numerous data checks to identify quality issues, including monitoring survey completion patterns and identifying "speedsters" and "straight-liners." We also work closely with panel companies who employ other tools, such as cross-checking respondent identity during panel recruitment, re-validating panel participants over time, and using digital fingerprinting to ensure a single IP address cannot be used more than once for the same survey. In short, the combination of a rigorous researcher and conscientious panel company could sufficiently mitigate the risk up until recently.

Wider industry recognizes the issue of survey fraud

However, there is a growing problem in one area we have far less control over – fake respondents and survey fraud coming from survey farms and "bots." And it is not just us flagging this problem, the MRS announced just a few days ago that it will be working

together with ESOMAR, The Insights Association and SampleCon to "track and address sample fraud and the risks it poses to data quality." As Jane Frost, Chair of the MRS mentioned "fraudulent activity is becoming increasingly sophisticated, particularly in online research," and in the same press release Melanie Courtright, CEO of the Insights Association said the issues are "global and persistent and must be addressed."

Bots and survey farms are becoming more sophisticated

Both myself and my colleagues, all of whom have 20+ years' experience in the industry, have never seen such serious sample quality issues as we are facing currently. The problem manifests itself in classic ways, such as inconsistencies in survey data. However, it is not a case of respondents straight-lining or taking surveys quickly – the bots operating within the survey farms can submit what look like valid response patterns in a reasonable (i.e., not too short) period. They can also answer all question formats and write what looks like, on first view, valid responses to open ended questions. For the latter of these, it is only when looking at the wider data set yourself when you can start to see similar looking responses using variants of the same set of words. These are not necessarily picked up by panel quality checks as they look valid on first view, so the onus has increased on researchers to establish new methods to spot them and remove them from the sample.

The extent of the problem

The survey farms are widespread enough and sophisticated enough to mimic legitimate respondents from multiple countries and can generate (one would assume), enough in the form of survey incentives to make it worthwhile for the perpetrators. For some of the online surveys we have conducted recently with a range of well-known panel providers, up to 20% of the respondents have clearly been generated by bots, and it is getting harder and harder to spot them.

So, what can we do about it? While there are welcome ongoing efforts from the industry mentioned above, we must take pragmatic steps while the industry (and technology) catches up.

Solution 1: Use a range of question formats, especially those requiring logic

- While "bots" can answer most questions in an online survey in what appears a valid way, they struggle with more complicated formats, such as "100% calculator" and "chip allocation" questions (where the respondent has to allocate proportions of a defined amount to different question options), as well as "drag and drop" and "slider" question formats.

- An additional quality measure is the use of "red herring" questions, where multiple answer options are provided including an obvious answer to a simple question. Also, trick questions in the survey that require a later survey response to match an earlier one can also prove helpful. Bots completing surveys in a "random" fashion may miss obvious or matching answers.

Solution 2: Consider more robust sample sources

- Alternative sample sources and recruitment methods can offer an even more effective solution. Poor (or clearly fraudulent) samples are more prevalent in our experience in B2B surveys (possibly because of survey farm operators targeting the more lucrative rewards available for completing these surveys). As such, we have found using a Phone to Web (P2W) approach is a viable solution in many cases for B2B surveys. Using this approach, our field team conducts an initial wave of phone recruitment to target and qualify potential survey participants and only after they are validated and screened is a link sent for the online survey. The difference in results vs. conventional online B2B sample (where we have trialed both sample sources before rolling out full fieldwork), has been startling – quite literally day and night. While the P2W method is more expensive than the online sample, the difference in quality and therefore confidence in the results more than makes up for the cost difference. Fake respondents and spurious data are of no use to anyone, regardless of the cost.

Looking ahead, it is encouraging to see steps being taken to address the problem, indeed some panel providers have already started to integrate new fraud detection tools. However, while we wait for this to become more mainstream, we strongly advocate using the measures ab