

Curbstoning and culture

Arthur Kennickell¹

Board of Governors of the Federal Reserve System, Washington, DC, USA

E-mail: Arthur.Kennickell@frb.gov

Abstract. This paper examines various aspects of data fabrication, or “curbstoning” in field surveys. Our ability to detect and control such behavior is limited by the costs of the most effective instruments, the weakness of most of our instruments and our limited understanding of what drives such behavior. Culture, an emergent pattern of thoughts and behaviors from a larger complex of systems and behaviors, operates in spaces where direct incentives or control cannot reach. Monitoring is important, but fostering a healthy culture among field staff may be the most efficient and the most humane approach to controlling curbstoning.

Keywords: Fabrication, field work

It is often helpful start by defining the universe of a problem, in this case “curbstoning” or survey data fabrication. At a highest level of abstraction, curbstoning might be described as active or passive behaviors by interviewers who knowingly deviate from the behavioral expectations of the survey managers. At a more practical level, I take curbstoning to be any kind of falsification by an interviewer of survey data or supporting paradata. This could be an interview that has been deliberately fabricated completely or in part. In the latter case, an interviewer might make up an answer for a respondent for a single question or a larger part of an interview where the respondent was a knowing or unknowing accomplice to the fabrication. Curbstoning could also be taken to encompass other actions that rest on inadequate effort, such as accepting unnecessary item non-response or unit non-response; such actions may also be accompanied by fabrication of paradata or

other data to justify the case not being completed or an answer not being collected. It could include many other types of distortion of paradata, including actions that exaggerate the amount of effort needed to complete a case. There could also be combinations of such behavior. These behaviors may have different motivations and means of detection, remedy or prevention.

Curbstoning is problematic for a variety of reasons. The most obvious cost, aside from the direct cost of paying an interviewer for work that was not performed, is the damage to the analytical integrity of a survey – except in the rare event that the interviewer has perfect knowledge about the situation of the respondent. For policy-driven surveys, a result of fabrication could be a misallocation or waste of scarce social resources. Sometimes there can even be harm directly to respondents. Kristin Himelein at the World Bank gave me a compelling example from a panel survey. In the baseline interview for one case in the survey, the interviewer collected information on the respondent, but fabricated the remaining family structure and associated data. When another interviewer returned to conduct the second wave using dependent interviewing involving information on the family structure in the baseline, the man’s wife was present and refused to believe that her husband was not also maintaining another family. An additional and very serious consequence of curbstoning is the potential for corruption or demotivation of other interviewers or other parts of the survey process if such behavior is not treated seriously.

¹Opinions expressed in this paper are my own views and they do not necessarily represent the views of the Federal Reserve Board or its staff. I would like to thank far more people than it is possible to thank in an article. The many survey field and support staff I have met in my career have generously opened their thoughts to me and shaped my thinking in profound ways about how people ought to behave and how they behave in fact. Catherine Haggerty, Ella Kemp, Sandra Pitzer and Nina Walker are only a very few among the many wonders for which I must be grateful. I also thank Fritz Scheuren, my dear friend and colleague for many decades, for asking me to give the talk upon which this paper is based.

Curbstoning is possible in any interviewer-mediated survey; but it is more difficult to address in field surveys than in a typical CATI survey, because of the more limited ability to observe or affect behaviors directly. Arguably, an even more difficult problem is the determination that curbstoning has occurred. Interviews do not come marked as fraudulent. Generally, a very substantial effort would be required to validate fully every interview, and other tools for detection tend to be weak or costly in some sense. One implication of the difficulty of observing curbstoning is that the measurement of seriousness of the problem in a given survey also may be limited. If curbstoning is rare, as it should be if a survey has any credibility, the cost of detecting instances of curbstoning may be very expensive per observation or interviewer identified; moreover, great care is necessary in applying any monitoring or evaluation process, to avoid complicating the work of the trustworthy great majority.

In order to think clearly about how to address this situation, it is important to consider the possible range of motives affecting the decision to fabricate data. In deciding to act in a particular way, interviewers must weigh the expected external and internal benefits and costs to them of what they are expected to do. That balance might be affected in a variety of ways and some aspects may be open to some degree to the actions of survey managers. An interviewer might be led to falsify data by disaffection, alienation, desperation, greed, misinformation about what is important for the survey, psychological comfort or discomfort, or other personal needs or preferences. Factors arguing to an interviewer against such actions might include fear of being detected and the related consequences or various internalized beliefs or values.

An interviewer who lacks any positive external reason to act as instructed – either because project-specific training was inadequate or because the interviewer did not find the training credible – might deliberately falsify an entire interview simply so as to be paid for minimal effort or sometimes to appropriate incentives intended for respondents. In arguably the worst situation, an interviewer might actively seek to subvert the goals of a survey, because of dislike of management, the goals of the survey, or some other aspect of social order – or possibly as a result of more personal antipathies.

In part because the most critical work of interviewers is so difficult to observe and reward directly, their performance is often evaluated on such factors as the number of completed cases and the associated costs.

Such pressure sometimes may distort the emphasis that they place on the unobserved aspects of their work, and in extreme cases may lead them deliberately to cut corners and even fabricate data. Several types of misinformation may also contribute to such behavior. Often, there is a misperception that survey statisticians are able to deal well enough with virtually any sort of problem, as long as they have “enough” completed interviews, however they are completed; an interviewer who had such beliefs would tend to place too small a weight on the importance of following survey protocols. Obviously, interviewers who do not have a sufficient grasp of the survey protocols may also weight desired behavior inappropriately. Interviewers who have been given no strong reason to execute a survey or follow the survey protocols will have no motives to offset their own convenience or desires, except their own standards of appropriate behavior, which obviously are often considerable.

Sometimes surprising motives drive falsification. In another story related to me by Kristin Himelien, an interviewer who had been detected falsifying interviews revealed that the motive had been to avoid getting fleas from the tents where the survey subjects lived in the winter. Understanding motivations can sometimes lead to means of altering the balance for interviewers. For example, in this case, perhaps anti-flea spray for interviewers would have been a positive incentive to behave according to plan. Where possible, resolution of problems related to falsification should involve an examination of the motives that drove the interviewer and that information should be fed back into the design of future survey work.

Partial falsification may be driven by similar motives; but such behavior tends to be far harder to detect. It seems likely that the dominant factor may be the discomfort of asking certain questions or the fear of “losing” the entire interview to a breakoff.

A typical approach to quality control works poorly for curbstoning, because our tools for detecting it tend to be costly and relatively weak and because curbstoning is a rare event – just as most people are not criminals, most interviewers do not falsify interviews.¹ Field interviewers are difficult or costly to observe directly in the course of their work, and intervention at that point might also risk weakening both the commitment of the good interviewers and the trust respondents need

¹In my opinion, where there is any significant number of interviewers fabricating data, no survey should be conducted until a credible remediation plan can be developed.

to have in the confidentiality of a survey. One common method of searching for complete fabrication is recontact of a sample of each interviewer's respondents. Where a sign appears that an interview may have been falsified, typically all or a higher fraction of the interviewer's cases would be recontacted. This approach is labor intensive, and it can be inconclusive, because respondents are sometimes surprisingly confused or forgetful about what did occur, because an interviewer may ask some small set of questions to "seed" a largely fabricated interview, because contact is often difficult and the information necessary for recontact is subject to manipulation by interviewers, and because of other factors. There have also been instances of collusion between interviewers and respondents.

Close examination of the survey data and paradata, as has historically been the case for editing in the Survey of Consumer Finances sponsored by the Federal Reserve Board, can also provide insights about the relative likelihood of falsification; but the effort is very costly and data anomalies that might signal curbstoning might well also be a more routine sort of nonsampling error. It is at least possible that paradata on contact attempts might be more amenable to mechanical manipulation to identify candidates for closer examination; but obviously there are also many possibilities that could be missed there as well. Organizing investigations by interviewer might help in identifying large-scale offenders, but would be unlikely to be of much use in dealing with an interviewer who fabricated data relatively infrequently.

Some relatively light technological interventions, such as using GPS devices to ensure interviewers are in the locations they say they are in or recording of random snippets of interviews using the microphone of a laptop computer used in a CAPI interview, might be helpful and manageable. But until more intelligent software is available, review of that information would need to be performed by some person, and even then the information content about what was actually happening could still be subverted by someone who understood what was being done.

The survey managers are a critical element in many aspects of data quality, particularly in listening to interviewers' problems and coaching them to a higher level of performance. Interviewers who have problems with authority or following rules are relatively likely to be noticed at this level. In reviewing each interviewer's cases regularly, the managers also have the opportunity to notice anomalies in what may appear to be unexceptional completed cases.

Making known to field staff that survey data are monitored for fabrication, and addressing any detected instances very seriously may be helpful in deterring some people from such behavior. But with our weak detection abilities and our limited ability to observe or control behavior directly, a great deal depends on the thoughts and belief structures of individual interviewers and the social structure that surrounds their work. In my experience, nearly every interviewer has been sufficiently committed to the work that fabrication is not a concern; for that reason, one should be very wary of instituting procedures that damage the sense of commitment, motivation, fairness, trust, etc., felt by them while addressing a problem created by a small minority.

As an economist, I feel obliged to introduce the element of compensation as a possible factor. Interviewers are paid for their time working and it is conceivable that we might pay interviewers enough more than the prevailing wage that they would develop a similarly stronger desire to follow the rules so as not to risk losing the job. But again, because the ability to monitor is so weak, it is generally not possible to fine tune payments to encourage desirable behavior directly. Any positive effects from a given level of payment may operate in part through a channel that supports interviewers' self-respect and perceptions of respect by others.

Problems similar in some respects to curbstoning exist in all of the parts of an interviewer's work that are not directly observed, for example the area of protecting confidentiality. For confidentiality, we tend to emphasize legal protections and penalties for violations to get interviewers' attention; but much effort is devoted to getting interviewers to recognize the central importance of protecting confidentiality and integrating it deeply into their thoughts and practices – what is often called creating a "culture of confidentiality." In a healthy survey organization, protecting confidentiality is taken as a given, with exceptions being viewed in that social culture as a high crime and unthinkable for virtually everyone. Values that sustain this positive behavior are internalized and are reinforced by social norms and expression of those norms is taken as an indication of belonging in that world. There are other "cultural" structures that affect the loosely observed part of interviewers' work, such as their adherence to protocols aimed at data quality.

Culture may be seen as an emergent pattern of thoughts and behaviors from a larger complex of systems and behaviors. Developing and sustaining any type of healthy organizational culture requires serious

commitment. But where monitoring is impossible (or very weak and costly), culture is the only vehicle available for influencing behavior. Where the difficult-to-observe behavior is also a critical element of work, as it obviously is in the case of field interviewers, the importance is further amplified. For now, there is no realistic alternative for addressing curbstoning.

A foundation of culture is a reasonably congruent understanding across a group and a sense of shared identity, purpose and values. Training can be effective in giving people a common set of facts; but facts alone are unlikely to be enough to yield understanding. Understanding is also often more of a systemic quality that encompasses internal representations of facts, context, relationships and the co-movement of those factors over time. In some ways, teaching for understanding is akin to teaching someone speak a language or to dance. Training is a start and continual reinforcement and elimination of misconceptions can develop understanding and keep it alive. For interviewers, it is obviously important to cover the relatively concrete “rules” of behavior; but it is also particularly important to communicate as clearly as possible the broader implications of a survey and the effects of interviewers’ behavior on events often far beyond them.

Identity, purpose and values can flow from and govern the application of understanding. Specifically recognizing the importance of interviewers may serve to heighten the common identity and purpose and to encourage feelings of responsibility. This must be more than simply being nice to people – it is a matter of sincere and deep respect presented and acted on at every available chance. Positive peer pressure may also emerge if some people appear to be violating what becomes a social norm. This factor may operate relatively indirectly in most surveys, since interviewers’ work is typically solitary. However, some surveys, such as the Chinese Household Finance Survey run by Li Gan and his colleagues at Southwest University for Finance and Economics, use teams of interviewers, in this case teams of students, who encourage and effectively monitor each other.

I am uncomfortable writing about the possibilities for interviewers to misbehave. In my experience, interviewers are almost always the unacknowledged heroes of a survey. Aside from the interested motives of a project, justice calls for recognizing these critical participants in social measurement. To the extent that interviewers who feel valued also act to protect the system that values them, so much the better.