May DOUŠAK, Roberto BRICENO-ROSAS, Joost KAPPELHOF*

VIRTUAL SURROUNDING IMPRESSION: ETHICAL AND PRIVACY-RESPECTING TRACKING OF FACE-TO-FACE COMPUTER-ASSISTED PERSONAL INTERVIEW LOCATION**

Abstract. *Conducting high-quality, face-to-face, computer-assisted personal interviews is a demanding task that often calls for experienced interviewers holding specialised expertise and the motivation to gather the best possible survey data. Unfortunately, some interviewers may attempt to navigate the challenges of the surveying by resorting to different types of undesirable interviewer behaviour ranging from minor infractions like speeding to more serious transgressions such as partial or even complete data fabrication. Given that even a small number of fabricated interviews can contaminate an entire dataset, it is imperative to swiftly identify the most serious forms of undesirable interviewer behaviour. With ethical and privacy considerations of interviewers and respondents in mind, we developed a novel approach called the Virtual Surrounding Impression (VSI), which allows the gravest forms of undesirable interviewer behaviour to be detected without resorting to actually recording location, audio or video data while taking account of ethical and privacy concerns. We show that the VSI approach enables the detection of instances where a single interviewer conducts multiple interviews at the same location or where the location changes during an interview, signalling possible fabrication, whether partlialy or full.*

Keywords: *survey interview fabrication detection, undesired interviewer behaviour, survey interview location, virtual surrounding impression.*

* May Doušak, PhD, teaching assistant and researcher, Faculty of Social Sciences, University of Ljubljana, Slovenia, e-mail: may.dousak@fdv.uni-lj.si; Roberto Briceno-Rosas, researcher at GESIS– Leibniz-Institut für Sozialwissenschaften, Mannheim, Germany; Joost Kappehof, PhD, head of the Methodology Department at The Netherlands Institute for Social Research (SCP), Netherlands. The authors are members or the European Social Survey Core Scientific team and working on the Undesirable Interviewer Behaviour detection package.

** Research article.
DOI: 10.51936/tip.61.3.669

INTRODUCTION

Surveys are used to collect data in almost all fields of social science research. Data gathered in this way are used daily by countless academics and professionals, journalists and even the general public, which uses it to make informed decisions. On the basis of surveys, data policies are made, products are introduced, changed or withdrawn, communication with the public is carefully fine-tuned, while countries may even use surveys to rank themselves according to a broad set of indicators.

The evident impact of survey data means that its quality should be beyond question. The survey research community has learned a lot since the Literary Digest debacle of 1936. Today, there is, for example, a far better understanding of how people read and answer questions, how to devise a questionnaire, the importance of probability-based samples and the impact of the survey interview setting (Doušak 2017; Squire 1988). In short, we understand quite well how to obtain and deliver high-quality survey data.

Despite this vast body of knowledge showing us how to obtain accurate survey data, other important considerations also play a role that can impact the quality of survey data. For example, such data quite often need to be delivered in a timely manner, preferably at a low cost. This might influence survey-related decisions made by a survey agency or individual interviewer. When fieldwork progress is slow and pay depends on the timely delivery of interviews, an interviewer might choose to behave in undesired ways, from milder deviations such as speeding and tacitly reacting to responses to, more seriously, partial or complete data fabrication.

The discussion concerning the morality of survey implementers and data fabrication is not new. Already in 1945, when discussing interviewers' work Crespi observed that "*human beings, it seems, remain human beings when they are asking questions for polling*". Further, he identified some "demoralizers" (incentives to cheat) and methods for detecting cheaters and suggested several methodological as well as administrative changes to reduce the risk of cheating.

Many statistical methods of detecting fraud in surveys have since been developed since and it is now possible to at least roughly estimate the proportion of fabricated data in a survey. Studies generally suggest that fabricated cases in large-scale surveys in OECD countries rarely exceed 5%, although up to (and above) 25% is not unheard of in some studies in non-OECD countries (Koch 1995; Kuriakose and Robbins 2016; Li et al. 2011). There are multiple reasons for such a huge discrepancy: from a lack of funding, an absence of effective quality monitoring during fieldwork, no quality assessment procedures prior to releasing the survey data to the public and/or creative approaches taken by the survey agencies or national coordinators to primary sampling units living within dangerous neighbourhoods for which the interviewers fabricate data to avoid having to enter them (Finn and Ranchhod 2015).

It seems that Crespi was correct 80 years ago when stating that

> *almost every interviewer will eventually succumb if the incitements to fabrication are made overpowering enough, if fabrication is made to appear the only practicable solution to the problems facing the interview.* (Crespi 1945)

Nevertheless, high-quality large-scale projects check survey data rigorously before publishing and may decide not to publish data in the event of data quality issues. For example, cases were removed from the European Social Survey dataset due to quality issues and an entire country's dataset was even barred[1] from inclusion in the integrated data file altogether. In such instances, both the costs incurred (interviews were paid for) and the reputation (the excluded country is explicitly mentioned in the ESS Data portal) may incentivise the closer monitoring of fieldwork (ESS 2024a).

Still, and despite the shrinking share of the survey data collection market in the West and new self-administered data collection modes offering advantages like faster data collection processes, a greater feeling of privacy while responding, more effective fieldwork administration, new presentation possibilities and often reduced costs compared to traditional face-to-face surveys, face-to-face interviewing quite often remains the only option for collecting valuable data about people. For example, when the population under study is unfamiliar with survey research an interviewer can help persuade and administer the questionnaire, thereby reducing the chance of refusal, incorrect answers, or misunderstanding. They can also be more effective in keeping people motivated to finish a survey compared to self-administered surveys (Uhan 1999). Moreover, a wider demographic can be reached via face-to-face interviews compared to most other modes. Notably, in many countries there remains a substantial portion of the population that is unable to conduct a self-completion survey due to being functionally illiterate. All in all, face-to-face still provides comparatively higher response rates than other survey modes (Daikeler, Bošnjak, and Lozar Manfreda 2020).

In addition, countries sometimes lack the infrastructure to conduct a survey without the use of interviewers, for instance when an interviewer is required to perform vital tasks such as selecting the right person to be interviewed when no useful sample frame is available. To avoid this problem, researchers may sometimes resort to non-probability web panels, notwithstanding that the research advises against doing so (Vehovar and Čehovin 2022).

Hence, despite the decreasing willingness to participate in any survey mode and professional interviewers becoming very rare in some countries following the coronavirus pandemic, face-to-face is still a valid (or the sole) option in some cases, which means improvements of the mode remain necessary.

[1] Data from Austria were excluded in rounds 4 and 5, data from Lithuania in round 4, and data from Albania in round 9 (ESS 2024a).

IMPROVING FACE-TO-FACE DATA COLLECTION

Like all other modes, face-to-face data collection is evolving methodologically and technologically. Upgrading interviewer-administered surveys from paper and pencil (PAPI) to computer-assisted personal interviewing (CAPI) has brought considerable benefits in terms of data quality: the computer automatically takes the locus of control and routes through the questionnaire. It also allows the recording of new types of paradata such as timestamps, while data can be transferred to a central location directly after the interview, making continuous fieldwork and data quality monitoring possible. All this allows for easier and more effective fieldwork coordination with potential for monitoring individual fieldworkers' performance and the near-instant remote reassignment of cases.

Further, timestamps (the record of the exact time and duration of an interview, a section of the questionnaire, or the time it takes to ask and answer a single question) not only allow for the detection of unlikely interview duration (speeding or fabricating at a rapid pace), but also an unlikely interview time (late-night hours), short delays between multiple interviews by the same interviewer, or even multiple concurrent interviews. Unorderly timing also suggests data were altered after being collected, while gaps in the timing suggest multiple sittings. Deviations can be explained due to the interview situation[2], software glitches, and programming and coding errors, but when such patterns repeat in multiple cases of a single interview they should be carefully assessed, and the interviewer's work may need to be back-checked.

Continuous monitoring of progress and work patterns is critical because it also enables the detection and correction of specific problems in advance before the fieldwork is completed and the required net effective sample size is not reached or statistical analysis suggests the final dataset is poisoned with bad data that has to be removed.

Besides timestamps, interviewers and their work can be monitored using different paradata and approaches. For example, some fieldwork agencies may ask their interviewers to record the GPS location at the respondent's doorstep as proof of the visit and the basis for reimbursing travel costs, while others even record video or audio during an interview[3]. Such a practice requires an appropriate legal basis (contracts) and is unusual and often undesired from an ethical

[2] While all undesired, they are not necessarily an indicator of an interviewer trying to corrupt the data collection process. To name a few, speeding can be explained by the respondent being on the verge of breaking off due to the questionnaire length (meaning the interview was sped up to prevent a break-off); some respondents prefer early morning or late evening interviews; short delays can be explained by two (or even more) assigned respondents being neighbours; an interview might have had to be stopped and ended at a later time due to unforeseen circumstances; the respondent might have changed their mind and amended past questions; and finally, the computer clock can misfunction or automatically readjust during an interview.

[3] Interviewers and even representatives of some survey agencies usually reveal their (quite comprehensive) paradata collection during informal talks, but rarely report it formally and even less so in peer-reviewed journal articles. Nevertheless, audio, photo, video and GPS recordings have already been researched and written about and may be consulted by the interested reader (Robbins 2018).

point of view if both the respondent and interviewer are not informed. Other, more indirect methods of data quality assessments may involve assessing item nonresponse or item non-differentiation as an indicator of interviewer behaviour. These indirect methods are usually more informative when more than a single interview is conducted by the given interviewer, while the former, such as GPS location logs[4] and audio recordings, offer evidence concerning the quality of the work for each individual unit of response. Aiming to combine the strengths of both '*direct*' and '*indirect*' monitoring methods, the authors decided to develop a new way of detecting the possible misconduct of interviewers.

THE EUROPEAN SOCIAL SURVEY

The European Social Survey (ESS) is an academically driven cross-national survey conducted across Europe since having been established in 2001 (ESS 2017). Every 2 years, face-to-face interviews are conducted with newly selected cross-sectional samples. This is considered by many to be the golden standard of an internationally comparative survey. The ESS has always been at the forefront of methodological developments and contributed to advances in survey data collection. For example, while many of the participating countries used computer-assisted personal interviewing before, CAPI with automatic time-stamp recording has been required ever since round 8 in 2016. Participating countries have been required to run their interim data through an interim dataset quality assessment tool since round 10 and the ESS is currently working to harmonise the data collection software in all countries in the upcoming round 12, scheduled to run in 2025–2026 (ESS 2019). Mindful that the detection of undesirable interviewer behaviour requires time and effort from every party involved in the quality monitoring, the ESS does not simply set requirements, but also makes additional tools and assistance available to member states to use and assess the quality of their data at any time without extensive knowledge of the statistical methods for detecting survey fraud. Between ESS rounds 9 and 11, the ESS developed and improved a set of tools that allows parties such as national coordinators (a national representative or team, usually a prominent social scientist or university group, responsible for implementing the survey at the national level) and survey agencies to run locally on their national ESS data to assess the quality of the fieldwork and identify problematic interviewers. To name just a few aspects, the tool provides a PDF report with information on item non-differentiation (detection of near-duplicates), unrealistic times of interviewing or delay between individual cases by the same interviewer, unlikely interviewing speed and others. All of this is done with great consideration for ethics and privacy, taking the GDPR and national privacy laws into account.

The ESS treats its interviewers and (potential) respondents with the utmost respect. While the ESS wants to provide its users with data of the highest possible

[4] The authors are aware of »curb stoning«.

quality, it values their data providers' privacy and does not record them in any way, or even track their location.

Therefore, for round 11 we focused on developing a novel way of detecting interview fabrication: we successfully developed a privacy-respecting indicator that alerts the agency when an interviewer fabricates multiple interviews at the same place. This is achieved without recording any location data but by inferring the interview surroundings from other data instead. Given that no physical location data points are recorded or saved at any stage, we call it a *Virtual Surrounding Impression* (VSI).

RECORDING INTERVIEW SURROUNDINGS: THE ESS EXAMPLE

In the ESS, interviewers generally receive batches of sampled persons (up to 48[5]) they need to visit and interview at their homes, which means cases by the same interviewer are typically conducted in different locations. Sometimes, the interviews are conducted in a public space such as a café, making multiple interviews taken at the same location by the same interviewer possible, but the same location may often indicate suspicious activities like interview fabrication.

Technically, it is possible to check if an interview happened at a given location using GPS. Here, one would record the location data (e.g., GPS coordinates) of all interviews and later check whether the coordinates match the respondents' addresses. Alternatively, the agency might check if (approximately) the same locations are reoccurring for the same interviewer, suggesting the interviewer conducted multiple in-house interviews at the same location, which is unlikely.

In any case, the ESS believes that recording GPS coordinates (even in an encrypted or obfuscated form) amounts to a significant and unacceptable privacy intrusion, which led us to develop an ethical and privacy-respecting alternative.

Along with the privacy requirements, the key criteria while developing the indicator were:

i. make it ethical and privacy-respecting;
ii. never save any location data and make it impossible to infer the actual location from the saved data;
iii. make it impossible to replicate the location code by another member of the (national) interviewer team or the supervisor simply revisiting the location;
iv. the solution should work on most computer devices without any additional peripherals (like GPS);
v. the solution should be portable and easy to implement on any computer platform, using any survey software; and
vi. it should be free[6] and open source software, available to all interested users.

[5] 48 as per the ESS specification, although that number can slightly increase in some cases if approved by the ESS (ESS 2024b).

[6] To use Stallman's analogy, the solution is both *free as free speech* (*»libre«*) and *free as beer* (*»gratis«*) (Stallman 2015).

Point ii) was the most difficult criterion to meet: how can one record location information without storing the location data even in the computer's volatile memory? In theory, GPS coordinates could be recorded and then irreversibly encrypted into a »location code«, although that approach would oppose the first criterion as the actual physical location data are stored in a computer before the encryption. Moreover, the device would need a GPS tracker, thereby also opposing point iii).

Our solution approaches the problem from a different angle. Instead of recording (and then possibly scrambling) actual location data, unique information about the interview surrounding is detected, then irreversibly transformed and, finally, recorded in the form of a fingerprint. The final impression is not constructed from physical location data (hence »*virtual*«) but works by detecting the *surroundings* and is stored as a short *impression* of the location.

Most portable and wearable devices in any form or size today include an IEEE 802.11 WLAN controller[7] and the networks they use are ubiquitous around the world. Given this, the next best data source of interest was data concerning nearby publicly broadcasted (visible)[8] Wi-Fi access points. Visible access points publicly broadcast information on the networks they provide such that any nearby device can automatically connect to them provided it has sufficient authentication information, such as a network password or client certificate (IEEE 2021).

To make things more convenient for the user, most client devices such as tablets, laptops and phones automatically connect to any previously established network without any user interaction. This is only possible because the device periodically scans the Wi-Fi surrounding by default to check for the presence and signal strength of a known network. Thus, if the indicator relies on the Wi-Fi surrounding no *active* monitoring is needed since the list of nearby networks is available from the device's operating system at all times if the Wi-Fi is not disabled.

Aside from other information, visible access points broadcast their network names as set by the users (SSID – Service Set ID) and hexadecimal addresses set by the factory (BSSID – Basic Service ID). The latter are uniquely identifiable (no two devices have the same BSSID) and hence it is a good starting point for the VSI approach: when the same Wi-Fi access point is present during multiple interviews it is most likely they were completed at approximately the same location. With more than a single visible access point within range, the location can be pinpointed even more precisely by triangulating the location from the BSSIDs and their respective signal strengths and accordingly this is what we record in the VSI.

Based on testing in densely populated areas, information on up to five nearby access points was an optimal compromise between accuracy and the amount of recorded data.

[7] IEEE 802.11 Wireless Local Area Network (WLAN) protocols are branded as »Wi-Fi«.

[8] Visible (non-hidden) networks should not be confused with »Public« networks that anyone can connect to.

Imagine an interview conducted in a sparsely populated rural area. If the respondent has Wi-Fi installed at their location and no other networks are in range, only their Wi-Fi is detected. However, in a more densely populated area with multiple Wi-Fi network points installed nearby it may happen that the same networks are detected at different locations (e.g., apartments). By also using signal strength, we can prevent such 'false positives'. Further, by recording up to five data points an interviewer using a mobile Wi-Fi hot-spot access point is not flagged as problematic' when the other surroundings change due to a change of the actual location.

Table 1 lists the five strongest visible access point BSSIDs at one of the author's offices:

Table 1: AN EXAMPLE OF ACCESS POINT BSSID AND SIGNAL STRENGTH DATA

14:16:9D:A8:5A:81	85 %
14:16:9D:A8:5A:8D	62 %
74:AC:B9:B1:E3:59	57 %
C0:C1:C0:0C:8E:97	42 %
64:66:B3:3F:B2:9A	40 %

Author's Data.

In terms of privacy, simply recording network data (base station IDs and possibly also signal strength) does not suffice: by revisiting the location, one might obtain the same surrounding data as the interviewer (opposing criterion ii)).

Moreover, although such a practice is questionable from a legal standpoint in the EU, there are software solutions like Google Geolocation API that translate Wi-Fi BSSIDs into actual location data (Burdon and McKillop 2014; Google 2024; Wang et al. 2017; Watts, Brunger, and Shires 2011).

MAKING SURROUNDING IMPRESSIONS FROM ACCESS POINT DATA: INTRODUCING A ONE-WAY FUNCTION

To prevent the location from being »decoded«, actual access point data (BSSID, signal strength) cannot be saved since anyone could use the Wi-Fi network information and convert it back into an actual physical location using a service like Google Geolocation API. This means the data must be obfuscated using a one-way data scrambling procedure which always produces the same output data for the same input and different output data for different inputs (AP information) while making it impossible to decode the final output back into the original information. In short, the data should be *hashed*[9].

[9] The principles of the inner workings of hash functions are beyond the scope of this article. There is, however, ample literature on the matter, such as Preneel (1994), which the interested reader can consult.

Using xxhsum[10], all the data from Table 1 can be hashed into a single hexadecimal string: d87e51daca20faa5. There is, however, a problem with this approach: if only a single digit is changed, like if the strength of the most powerful network increases from 85% to 86%, the output hash is completely different: 108e3527e6b8e8b4.

Hashing the Wi-Fi data as a whole, obviously, obfuscates the information too much, making the same surrounding appear different when only a single access point's signal strength changes slightly.

To solve this problem and enable the detection of the same surroundings, each individual access point's data point is hashed separately. With 5 access points, there are 10 strings to be hashed: 5 BSSIDs and their 5 respective signal strengths. The data shown in Table 1 can therefore be written as a single string containing 10 hashes (bold is used for legibility reasons for every second data point), which is quite easy to store as a text survey variable: ee4c31da5fec9c93 **a2a42b48293c27e4** dd9ea46415e7fa95 **471e21bba092b8a** 0345c98b3c080083e **291efdbbb3c8b971** 51a4e0aa42866ee5 **589a0f47037e35c01** 6104ad3de7ea8c6 **ff466310d44a539d**

Given that hash function outputs have predetermined lengths, the above text can be split back into individual hashes when performing analysis. This irreversible »code« of the surrounding data cannot be converted back into a physical location and thus we call it a *surrounding impression* because it deals with the problem as described above.

MAKING THE SURROUNDING IMPRESSION VIRTUAL

While it is impossible to convert the surrounding impression back into a physical location by means of a database like Google's Geolocation, theoretically anyone can obtain the same surrounding impression by simply revisiting the same location.

Further, while it is impossible to decode the *surrounding impression* back into the original access point data, one could create a »dictionary« containing all possible combinations of BSSIDs and signal strengths from which they could »translate« the hash back into the original data. Such dictionaries (called *rainbow tables*) are widely used in cryptoanalysis, yet they have practical limits when it comes to decoding codes of longer original data (Horálek et al. 2017).

In order to prevent the use of rainbow tables, a long and random string (called a »salt«) can be added to access point data before hashing in a process called *salting* (McGiffen 2022). Finally, to make the final code unique to each interviewer, thereby preventing the recreation of codes by simply revisiting the location, the *salt* must be unique to each fieldworker so that only their devices, and only when used by them, generate the same codes at the same locations.

[10] Xxhsum is a non-cryptographic open source hashing algorithm. For greater security, cryptographic hashing algorithms (e.g., the *SHA* family) can be considered for a production environment (Collet 2023; Dang 2015; Estébanez et al. 2014).

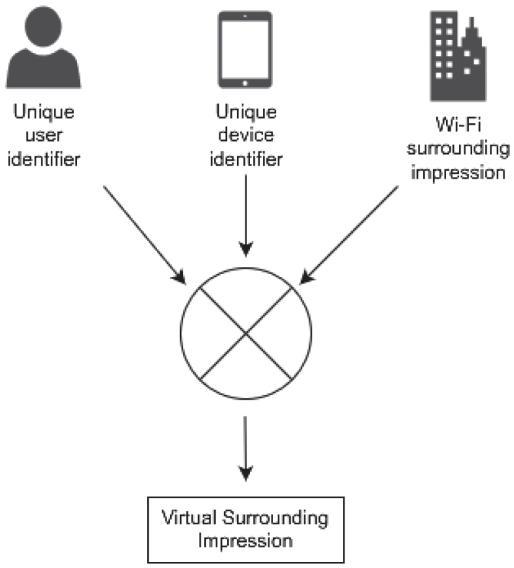
In most modern operating systems (OS), both the *device* and *user* have their own 128-bit long universally unique identifiers (UUIDs) that are generated when the OS is first installed or a user is created. When combined, these two identifiers provide a string that is unique for each interviewer and their device. In addition, using both UUIDs makes for a 256-bit long salt and prevents the use of rainbow tables.

Given that a salt is device- and user-specific, no other user or device can recreate the same *surrounding impression* at the same place, hence making the solution completely privacy-respecting, yet at the same time powerful enough to detect whether the *same user* using the same device was at the *same location*.

Construction of the VSI is quite similar to the construction of the surrounding impression. First, unique user and device IDs are obtained from the operating system so they can be used later as a salt. Then, each data point (each BSSID and its respective signal strength) is salted with a unique device and user ID and, finally, like with the prior *surrounding impression* each (salted) data point is hashed using a hashing algorithm of choice to produce the final hashes, which can be combined into a single string and stored as a survey variable.

In this case, the link between the original Wi-Fi surrounding and the final impression is completely broken, and we call the solution a *Virtual Surrounding Impression*. The concept is outlined in Figure 1 below:

Figure 1: CONCEPTUAL DRAWING OF A VIRTUAL SURROUNDING IMPRESSION



Author's Data.

On the condition that: (a) the virtual surrounding hashes never leave the survey agency; and (b) interviewers (as employees or subcontractors) are aware that *different technical means* are used to monitor their movements, the VSI was found to be ethical and privacy-respecting enough to be approved by both the ESS data protection officer and the ESS ERIC Research Ethics Board.

IMPLEMENTATION

The aim was to create a tool that is easy to implement on any operating system, without limiting ourselves to any existing survey software, hardware or software platform. While the tool was developed for desktop operating systems (laptop computers), it is released as a free and open source solution under the terms of copy-left GNU GPL version 2, which means developers are free to port it to any platform or alter it.

Given the intended usage, it was decided to record the VSI as an 80-character hexadecimal text string; users are free to use stronger cryptographic hash algorithms by a simple change of the command if they so desire.

Considerable care was given to make the solution simple and portable, such that it can be implemented on any operating system in a matter of minutes rather than hours. Accordingly, the choice was made to rely on command-line tools already implemented on all desktop (Linux, Mac, Windows) operating systems. An example implementation is included in Appendix A.

To make things easy for fieldwork agencies without in-house programmers, we created a Python script that detects the current operating system, executes the appropriate commands and returns a VSI code, allowing even administrators without any coding skills to implement it while installing survey software on interviewers' laptop computers. The tool is available from https://code.may.si/may/VirtualSurroundingImpression (Doušak 2023).

Data can be analysed using the ESS Interim Analysis tool, which also detects other types of data issues and undesired interview behaviour, such as a high workload, a suspicious time of interviewing, overlapping interviews, speeding, or item non-differentiation. The interim dataset analysis tool is free and open source and available from GitHub at https://github.com/briceno-rosas/ESS-Interim-Dataset.

REAL-WORLD TESTING: LESSONS LEARNED

The VSI was initially tested on a small scale in 2019, after which it was further refined: instead of continuously recording the data in a separate database every 10 minutes, data were recorded as a survey string variable three times during each interview; instead of data for 20 access points, 5 were used.

The pilot revealed that our signal strengths in *salted* and *hashed* form reveal nothing about the signal strength itself. As already described in this article, the final impression changes dramatically regardless of whether there is only a slight (1%) or significant (90%) difference in signal strength. As such, this information is of no use.

The signal strength can be stored in its original form (e.g., »90%«) without revealing much about the location as in any case the access point identification is encoded. Nevertheless, the choice was made not to record anything in non-encrypted form but to rely on the *order* of access points (based on the signal strength).

Further, practical testing revealed that in most cases simply checking for the presence of the same three (out of the five recorded) access points' BSSID hashes is a solid indicator of approximately the same location.

The virtual surrounding impression was introduced and fully tested in a real setting in three larger-scale surveys: Slovenian Public Opinion (*Slovensko Javno Mnenje*) 2019/2 and 2022/1, as well as the ESS Round 11 in Slovenia (conducted in 2023).

In those real settings, the VSI was used to detect whether:

a. the interviewer did not visit the potential respondent and completely fabricated the survey while at home;
b. the interviewer did visit the respondent, recorded their location, may or may not have briefly talked to the potential respondent (to gather information) and then fabricated responses at home; or
c. the interviewer visited the potential respondent and had them cooperate but skipped some blocks to make the survey shorter, before later filling in the blanks at home.

Still, the VSI cannot detect curb stoning, which refers to when an interviewer sits in front of the house of a potential respondent and completes the interview themself. In addition, all flagged cases were further evaluated using other means of quality control, primarily using time-stamp based analysis and back-checks.

Most of the flagged cases, which turned out to be deviations from the interviewing protocols, were also flagged by the time-stamp based analysis. The vast majority of those double-flagged (VSI and time-stamps) cases were minor deviations, such as later correction of coding mistakes, albeit there was also some minor partial fabrication.

A few cases that were flagged by the VSI also turned out to be false-positive after back-checking. Errors were attributed to Wi-Fi being switched off, surveys being conducted at a location without Wi-Fi or taken at a cafe, or technical issues[11].

Finally, one interviewer who worked for us for the first time had many cases flagged by the VSI, but none by time-stamp-based analysis. Comprehensive postal and in-person back-checks revealed that they had completely fabricated quite a few cases. This was an experienced interviewer who was aware of

[11] No method of detecting undesired interviewer behaviour is immune to false positives. To give a few examples from practice, time-stamp-based analysis can flag cases when the interviewer has gone back and forth, when the computer clock has issues (i.e., resets to an earlier time, resulting in overlapping interviews), the interviewer has opened the survey for the wrong respondent (recording an earlier or a later starting time), or when the interviewer has forgotten to properly end the survey.

time stamp recording and thus tricked the time-stamp analysis by fabricating it at a reasonable pace during a reasonable time. However, the interviewer was unaware[12] of the VSI and hence did not change location for each of the fabricated interviews and was caught because of the VSI indicator.

In its implementation, the VSI has proven to be an effective complement to time-stamp-based analysis. Like with most indicators aimed at capturing undesirable interviewer behaviour, we advise survey practitioners such as survey agencies or researchers to combine methods of undesirable interviewer behaviour detection and back-check flagged cases before taking any further action.

DISCUSSION

There are many known ways of fabricating data and other forms of undesirable interviewer behaviour (e.g., non-adherence to protocols set out in specifications), and describing them in detail lies beyond the scope of this article. Like most other techniques for detecting undesired interviewer behaviour, the VSI has limitations and cannot be used as a catch-all solution. Given the aims of this indicator, the focus was on detecting partial and complete interview fabrication by the interviewer.

To fabricate an entire interview, the interviewer needed a place to fill in the questionnaire. The ESS questionnaire takes roughly 1 hour, and most interviewers wanting to fabricate an interview are well aware of time stamp recording and therefore act accordingly while fabricating: they will complete the questionnaire at a »reasonable« pace at an »appropriate« time of the day (Ghirelli et al. 2022). They might fabricate from home while doing something else, like watching TV or completing other tasks, or even in a car or train while commuting. In any event, a single interviewer has a limited number of locations from which they might fabricate responses. The worst offenders might even have a routine of fabricating many responses at the same place while doing the same thing (e.g., every afternoon while supervising the kids).

The VSI cannot provide an actual interview location, which means it cannot be used for cross-checking data with a respondent's address. It can, however, disclose a *change* in location.

When a certain interviewer only fabricates a single survey response from their home, the VSI code for the given interview indicates a different location from the other interviews and hence the fabricated survey code looks legitimate since it differs from the rest. If, however, an interviewer fabricates more than a single response, it might be detected by the VSI. While, in theory, they could fabricate each response at a different location and avoid detection, time stamp recording limits offenders' options to places where they can stay for the time

[12] All interviewers were aware (and signed a contract) that their work was being monitored using different privacy-respecting technical means and that their work would be randomly back-checked. For obvious reasons, they were not aware of the inner workings of the virtual surrounding impression.

that an interview takes – approximately 1 hour for the European Social Survey. This makes it less and less appealing for an interviewer to fabricate an interview if they are aware of all the different ways of monitoring interviewer behaviour.

There are valid cases in which a single interviewer conducts multiple interviews at the same location. Some respondents might not allow interviewers into their homes, leading to the interview being conducted at a public location such as a restaurant or a university. Interviewers (and back-checked respondents!) should be able to explain such cases.

Further, the VSI can be recorded at multiple times during an interview. While this may sound redundant, it is concerning when the location changes during an interview. Such a change might be reasonable when an interview was conducted in two sittings at two locations. It is even more concerning if the VSI detects a change in location but there is no time gap in the timestamps. There could be technical explanations for such a discrepancy, yet it could very well also mean that the »interview« took place while commuting.

When flagged along with the gap in the timestamp recording, it may indicate post-interview alterations in the responses or other forms of undesired interview behaviour. It can also be detected when an interviewer visits a respondent to briefly obtain information from them and then fabricates the complete response from home since the recorded VSI (the interviewer's home) would then match all the other cases they have fabricated while at home.

There are three limitations to the current implementation of the VSI. First, it obviously cannot detect curb stoning. If an interviewer chooses to fabricate the interview near the respondent's home, similarly to a GPS location recording, the VSI cannot detect it as a deviation.

Second, the impressions will then all be the same if an interviewer disables the device's Wi-Fi functionality. This can be solved by adjusting computer permissions and preventing the interviewer from turning the Wi-Fi off. In addition, if a particular interviewer manually turns Wi-Fi off while interviewing while most other interviewers do not do so, this could indicate undesired interviewer behaviour.

Finally, the signal strength in a salted and hashed form reveals nothing about the signal strength itself and is therefore of no use in the current form. As a solution to the final shortcoming, we suggest either: (a) record the signal strength as a clean-text percentage. The 90% signal strength of »unknown« does not reveal the location any more than an »unknown« signal strength of »unknown«; or (b) skip the information on the power altogether and instead only rely on the order of access points (sorted by signal strength).

CONCLUSION

(Partial) fabrication of interviews is as old as in-person data collection itself and will probably be around for as long as we collect survey data (Crespi 1945). In order to deliver high-quality data, survey agencies rely on different fraud

detection methods, but aside from costly, timely and lengthy back-checking of all respondents, not many methods allow for effective direct monitoring of the fieldwork (as opposed to indirect statistical data analysis) in an ethical and privacy-respecting way.

By indicating the interview location without storing any location data, the VSI provides direct insight into fieldwork situations while simultaneously respecting privacy. The implementation proved easy and worked transparently, even on very weak laptops without any additional hardware[13], which may make the solution suitable for projects with very limited funding.

Practical testing proved the VSI to be a useful complement to time-stamp-based quality checking that can be run continuously while in the field to detect the gravest forms of undesirable interviewer behaviour as they happen, in turn giving time to reissue cases and deliver an uncontaminated dataset.

BIBLIOGRAPHY

Burdon, Mark, and Alissa McKillop. 2014. "The Google Street View Wi-Fi Scandal and Its Repercussions for Privacy Regulation". *Monash University Law Review* 39 (3): 702–38. https://doi.org/10.3316/informit.376209506308923.

Collet, Yann. 2023. "Xxhsum: Print or Check xxHash Non-Cryptographic Checksums | Xxhash Commands | Man Pages | ManKier". Xxhsum Man Page. https://www.mankier.com/1/xxhsum#.

Crespi, Leo P. 1945. "The Cheater Problem in Polling". *Public Opinion Quarterly* 9 (4): 431–45. https://doi.org/10.1086/265760.

Daikeler, Jessica, Michael Bošnjak, and Katja Lozar Manfreda. 2020. "Web Versus Other Survey Modes: An Updated and Extended Meta-Analysis Comparing Response Rates". J*ournal of Survey Statistics and Methodology* 8 (3): 513–39. https://doi.org/10.1093/jssam/smz008.

Dang, Quynh H. 2015. "Secure Hash Standard." NIST FIPS 180-4. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.FIPS.180-4.

Doušak, May. 2017. "Survey Mode as a Moderator of Context Effects". *Advances in Methodology & Statistics / Metodoloski Zvezki* 14 (1): 1–17.

Doušak, May. 2023. "Virtual Surrounding Impression Tool Source Code". May's Public Git Repos. 1 June 2023. https://code.may.si/may/VirtualSurroundingImpression.

ESS. 2017. "About European Social Survey (ESS)". http://www.europeansocialsurvey.org/about/.

ESS. 2019. "Methodology: Data Collection". European Social Survey (ESS). https://www.europeansocialsurvey.org/methodology/ess_methodology/data_collection.html.

ESS. 2024a. "Data Portal". https://www.europeansocialsurvey.org/data.

ESS. 2024b. "Survey Specification". http://europeansocialsurvey.org/methodology/ess-methodology/survey-specification.

Estébanez, César, Yago Saez, Gustavo Recio, and Pedro Isasi. 2014. "Performance of the Most Common Non-Cryptographic Hash Functions". *Software: Practice and Experience* 44 (6): 681–98. https://doi.org/10.1002/spe.2179.

Finn, Arden, and Vimal Ranchhod. 2015. "Genuine Fakes: The Prevalence and Implications

[13] AMD E-350 based Lenovo Thinkpad X120e from 2011 with 4 GB of RAM running a Debian Linux OS with an XFCE desktop.

of Data Fabrication in a Large South African Survey". *The World Bank Economic Review*, September, 31 (1): 129–57.

Ghirelli, Niccolo, Peter Lynn, Brita Dorer, Hannah Schwarz, Joost Kappelhof, Janine van de Maat, Georg Kessler, Roberto Briceno-Rosas, and L-M Rød. 2022. "ESS9 Overall Fieldwork and Data Quality Report". GESIS. http://europeansocialsurvey.org/sites/default/files/2024-03/ESS9_Quality_Report_20240326.pdf.

Google. 2024. "Geolocation API Overview". Google for Developers. https://developers.google.com/maps/documentation/geolocation/overview.

Horálek, Josef, Filip Holík, Oldřich Horák, Lukáš Petr, and Vladimir Sobeslav. 2017. "Analysis of the Use of Rainbow Tables to Break Hash". *Journal of Intelligent & Fuzzy Systems* 32 (2): 1523–34. https://doi.org/10.3233/JIFS-169147.

IEEE. 2021. "IEEE Std 802.11™-2020, IEEE Standard for Information Technology—Telecommunications and Information Exchange between Systems—Local and Metropolitan Area Networks—Specific Requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications".

Koch, Achim. 1995. "Gefälschte Interviews: Ergebnisse der Interviewerkontrolle beim ALLBUS 1994". *ZUMA Nachrichten* 19 (36): 89–105.

Kuriakose, Noble, and Michael Robbins. 2016. "Don't Get Duped: Fraud through Duplication in Public Opinion Surveys". *Statistical Journal of the IAOS* 32 (3): 283–91. https://doi.org/10.3233/SJI-160978.

Li, Jianzhu, J. Michael Brick, Bac Tran, and Phyllis Singer. 2011. "Using Statistical Models for Sample Design of a Reinterview Program". *Journal of Official Statistics* 27 (3): 433–50.

McGiffen, Matthew. 2022. "Hashing and Salting of Passwords". In Pro Encryption in SQL Server 2022: *Provide the Highest Level of Protection for Your Data*, edited by Matthew McGiffen, 269–75. Berkeley, CA: Apress.

Preneel, Bart. 1994. "Cryptographic Hash Functions". *European Transactions on Telecommunications* 5 (4): 431–48. https://doi.org/10.1002/ett.4460050406.

Robbins, Michael. 2018. "New Frontiers in Detecting Data Fabrication". In: *Advances in Comparative Survey Methods*, Timothy P. Johnson, Beth-Ellen Pennell, Ineke A.L. Stoop, and Brita Dorer (eds), 1st ed., 771–805. London: Wiley. https://doi.org/10.1002/9781118884997.ch36.

Squire, Peverill. 1988. "Why the 1936 Literary Digest Poll Failed". *Public Opinion Quarterly* 52 (1): 125–33. https://doi.org/10.1086/269085.

Stallman, Richard M. 2015. *Free Software, Free Society: Selected Essays of Richard M. Stallman*. Third edition. Boston, MA: Free Software Foundation, Inc.

Uhan, Samo. 1999. "Arbitrariness and Standardization of Survey Conversation". Teorija in Praksa 36 (5): 765–75.

Vehovar, Vasja, and Gregor Čehovin. 2022. "Izzivi uporabe neverjetnostih spletnih panelov v družboslovnem raziskovanju". *Teorija in Praksa* 59 (4): 849–70. https://doi.org/10.51936/tip.59.4.849–870.

Wang, Jin, Nicholas Tan, Jun Luo, and Sinno Jialin Pan. 2017. "WOLoc: WiFi-Only Outdoor Localization Using Crowdsensed Hotspot Labels". In *IEEE INFOCOM 2017 – IEEE Conference on Computer Communications*, 1–9. Atlanta, GA, USA. https://doi.org/10.1109/INFOCOM.2017.8057096.

Watts, Mark, James Brunger, and Kate Shires. 2011. "Do European Data Protection Laws Apply to the Collection of WiFi Network Data for Use in Geolocation Look-up Services?". *International Data Privacy Law* 1 (3): 149–60. https://doi.org/10.1093/idpl/ipr013.

Appendix A: Demo implementation

For Linux, we used nmcli for the WiFi list and machine-ID and current user's username sha1 code for salt, and then processed the data using standard command-line text processing utilities (sed, awk, tail, head).

Finally, we created a hash using xxhsum.

Linux one-liner solution[14]:

```
salt=$(cat /etc/machine-id)$(whoami | sha1sum | head –c 40) && for a in $(nmcli device wifi list | sed –e 's/^/*/g' –e 's/\w\s\w/_/g' | awk –v s=$salt '{print $2 s "\n" $7"%" s}'); do echo $a | xxhsum –H0; done | tail –n +3 | head –n 10 | cut –d ' ' –f 1 | xargs echo | sed 's/ //g'
```

Similarly, we used airport and ioreg commands on MacOS, and wmic and netsh[15] on Windows and embedded them into a Python script, which is available under GNU GPL license 2 at https://code.may.si/may/VirtualSurroundingImpression (Doušak 2023).

Demo implementation considerations

We strongly advise against sharing the same username and password over multiple fieldworkers and their devices (e.g., »fieldworker« user name on all machines). This not only creates the unique username part of salt the same for all fieldworkers in the VSI, but also makes the machines less secure against data theft.

While preparing multiple fieldworker machines, machine cloning may be the fastest and most preferred approach. After cloning, the machines' unique ids are the same, so the administrator must recreate them. This should also be done independently of the VSI, but is critical in terms of the privacy-respecting part of the VSI as machine ID is also used as part of the salt.

All Wi-Fi list commands (nmcli; airport; netsh) work passively without rescanning the network by default. Depending on the operating system defaults, the list of networks provided by these tools may be stale. If needed (and agreed with the privacy officer), rescan switches and appropriate system permissions to do so may be added to the commands.

Finally, given the intended use and the fact that VSI codes should never leave the fieldwork agency, the attack surface for a cyberattack is non-existent. Still, if one wanted, a stronger cryptographic hashing algorithm such as SHA instead of xxhsum could be used.

---

[14] Most Linux distributions do not provide a user's UUID. To make the solution universally applicable to any (POSIX) Linux, the above solution generates a sha1 hash from the currently logged in user's username and then takes the first 40 characters of that hash instead (whoami | sha1sum | head –c 40). Mac and Windows provide the user's UUID.

[15] Netsh wlan show networks only works when not connected to any network, which is reasonable when surveying at a respondent's home.

VIRTUAL SURROUNDING IMPRESSION: ETIČNA IN DO ZASEBNOSTI PRIJAZNA ZAZNAVA LOKACIJE PRI RAČUNALNIŠKO PODPRTEM OSEBNEM ANKETIRANJU

Povzetek. *Izvedba računalniško podprtega osebnega anketiranja na terenu je zahtevno opravilo, pri katerem je kakovost zbranih podatkov odvisna predvsem od izkušenih in ustrezno motiviranih terenskih anketarjev. Zaradi različnih razlogov se lahko posamezni sodelavci občasno zatečejo k bližnjicam, med katere spadajo različne vrste odstopanj od sicer visokih standardov anketarskega dela: od manjših nepravilnosti, kot je na primer hitenje skozi anketo, pa do najhujših kršitev, kot je ponarejanje anket. Že nizek delež ponarejenih anket lahko vpliva na celotno podatkovno zbirko, zato je ključno, da se najhujše oblike kršitev ugotovijo v najhitrejšem možnem času. Ob upoštevanju etičnih načel in varovanju zasebnosti anketarjev ter anketirancev smo zato razvili nov pristop, ki nam omogoča zaznavo najhujših oblik ponarejanja pri anketiranju. Z orodjem Virtual Surrounding Impression (VSI) lahko etično in brez shranjevanja lokacije, zvoka ali slike zaznamo spremembo lokacije pri anketiranju. S preizkusom na več raziskavah smo ugotovili, da orodje ustrezno opozori na anketarje, ki so izvedli ankete z različnimi respondenti na isti lokaciji, ter ankete, pri katerih se je med samo izvedbo spremenila lokacija, kar nedvoumno kaže na sum ponarejanja anket.*

Ključni pojmi: *zaznava ponarejanja anket, neželeno delo anketarjev, virtual surrounding impression, lokacija ankete.*