

Acquiescence Bias (Agreement Bias)

Simple definition: Tending to disproportionately select a positive response.

Full description: In particular, a participant is likely to give disproportionately positive responses to statements. It should be noted and therefore care taken that some cultures can provide more positive responses / aspirational responses.

Category: Behaviours and Biases

Aggregator

Simple definition: A supplier that provides access to participants by gathering multiple panel sources and making them all accessible via a single interface.

Full description: An organisation that provides access to participants by gathering multiple panel sources and making them all accessible via a single interface.

Category: Panels and Samples

Artificial Intelligence (AI)

Simple definition: Artificial Intelligence is a computing environment where the machine makes its own autonomous decisions.

Full description: Artificial Intelligence (AI) is a computing environment where the machine makes its own autonomous decisions and acts, creates, evolves, or changes decisions without the oversight or contribution of an actual human.

Category: General Terms

B2B Surveys

Simple definition: Business to Business research activities.

Full description: Business to Business surveys that are targeted toward business professionals. Examples of commonly targeted B2B groups include IT decision makers (ITDM), human resources decision makers (HRDM), and Healthcare providers (HCP). In addition to the types of fraud that threaten any research study, B2B surveys are particularly vulnerable to false claims of group membership that can severely compromise the validity of conclusions. Fraud rates in quantitative B2B studies are often much higher than in B2C studies, in part because of the higher monetary rewards associated with such studies.

Category: General Terms

B2C Surveys

Simple definition: Business to Consumer research activities.

Full description: Business to Consumer surveys that are targeted toward the general consumer.

Category: General Terms

Behavioural Validation

Simple definition: A process to identify problematic participants from their behaviour.

Full description: The process of identifying problematic participants through an examination of their behaviour, which can include responses to specific survey questions, response patterns across multiple questions, mouse movements and other behavioural measurement techniques. Behavioural validation can be applied both pre-survey and in-survey. This can be done at a survey level but also at a longitudinal level across multiple surveys over time. A participant may behave differently across surveys, being engaged in one but not in another.

Category: Tools and Methods

Benchmark Comparisons

Simple definition: Comparing data to a known baseline.

Full description: Benchmarks are generally used in order to assess how brands, products or services are performing in relation to a known measure - some research organisations have built up repositories of benchmarks (also known as "norms") over time that can be used to rate over or under performance. Any project measure, result or response that is outside of acceptable ranges versus available benchmarks and may suggest data fraud or quality issues. Care should be taken in assessing whether responses are likely to exist in reality.

Category: Tools and Methods

Blocklist

Simple definition: A list of entities that are denied access to a part of the online ecosystem.

Full description: A blocklist is a list of entities, such as users, websites, email addresses, IP addresses, or domains, that are explicitly denied access to a system, service, or network. The primary purpose of a blocklist is to prevent communication or interaction with identified entities that are considered undesirable, malicious, or unauthorized.

Category: Tools and Methods

Bots

Simple definition: Software that operates as an agent for a user or a programme or stimulates human activity.

Full description: Computer programs/scripts designed to mimic human activity to participate in online surveys for purposes the incentive or reward. Bots can be highly flexible with many designed behaviours. For example, they may be set up to pause during surveys to extend the length of time taken and avoid being flagged as speeding.

Category: Behaviours and Biases

Click Farm / "Organised Fraud Groups"

Simple definition: A group of people paid to complete surveys.

Full description: Significant activity from groups of participants generally for malicious purposes to participate in surveys and earn rewards. People are usually used to avoid bot detection technology. Surveys may be manually completed by people or they may use a bot as a tool to complete surveys at a higher speed than a human could achieve. Click farms can be a single person spoofing multiple accounts on one computer, a group of people working together, or a remotely distributed group working anywhere in the world. Click farms often use fake user profiles or automated scripts to simulate online activity. Click farms undermine the integrity of surveys by generating a large volume of fake and fraudulent responses. This can skew the data and lead to inaccurate findings. It becomes challenging for researchers to differentiate genuine responses from those generated by click farms. This leads to: biased responses due to not being able to accurately represent the target audience; not reflecting true opinions or user experiences; inflated response rates due to survey responses being artificially inflated. The result being that it is hard to determine true response rates and sample size.

Category: Behaviours and Biases

Contradictory Answers/Hidden Trap Question

Simple definition: An attention checking survey question.

Full description: Answers within one question, or across multiple questions where the data does not align. When a participant is unengaged or inattentive, or is showing acquiescence bias, they are likely to give contradictory answers. For example, if the statements "I like dogs" and "I dislike dogs" are put in different parts of the same list, you would expect them to get opposing responses if a participant is taking care to read the questions. If they give the same or very similar answers, it would be considered contradictory and would be considered a failure at the this hidden trap question. This can provide false positives e.g. someone could honestly like some dogs and dislike others and therefore be confused so

practitioners should always use multiple questions to identify the truly engaged from the unengaged.

Category: Tools and Methods

Data Fraud

Simple definition: Intentional misrepresentation of identity or data.

Full description: Deception or malicious behaviour intended for financial gain with fake insights provided or through system manipulation. Pre or during research or panel sign-up process, inaccurate or false information may be provided or participant flow system manipulated to register a person as qualified for research, most likely for the purpose of claiming an incentive. The term "Fraud" should only be used in the context of deliberate or malicious deception and should not be confused with poor "Data Quality".

Category: General Terms

Data Quality

Simple definition: The measure of the condition of data based on factors such as accuracy, completeness, consistency, reliability and how up to date it is.

Full description: There are a number of ways to check and measure data quality. In the research sector, it is critical for all parties to ensure that the highest possible quality data is collected, processed and analysed. For online data this might include checking open ended text answers, looking for patterns in data and so on. For other methods, it may involve listening to live calls or recontacting participants after a research task to verify qualifying criteria. The quality of data also includes other key aspects of research e.g, coverage / representivity - online does not readily cover the population and this can impact results. To ensure the highest possible data, it is important to follow best practices with regards to sample design, quota design, questionnaire design.

Category: General Terms

Digital Device Fingerprinting

Simple definition: Information collected about a device for the purpose of identification of individual research participants or devices.

Full description: Browser fingerprinting is a technique used to gather information about a web browser's configuration and settings to create a unique identifier or "fingerprint" for that particular browser. It involves collecting various data points, such as the browser version, operating system, screen resolution, installed plugins, fonts, time zone, language preferences, and other attributes that can be easily obtained through standard web technologies like JavaScript. By combining these data points, websites and online services can generate a unique identifier that can be used to track and distinguish individual browsers, even if cookies are cleared or disabled. The fingerprint is typically a combination of several characteristics, making it difficult for users to change or manipulate all the attributes to avoid tracking. In

research, browser fingerprinting can be used to ensure data quality in online surveys by identifying and mitigating fraudulent or suspicious responses. Browser fingerprinting can help identify and flag duplicate responses from the same browser or device. It is important to note that while browser fingerprinting can help detect fraudulent participants, it is not infallible. Dedicated fraudsters can find ways to manipulate or alter their fingerprints. Also, new browser versions, such as Firefox, offer ways to block fingerprinting.

Category: Tools and Methods

Disengaged Participant

Simple definition: A participant who reactively does not give an adequate level of thought to the responses they provide.

Full description: A participant who is disengaged and responds to questions without fully processing what is being asked due to several reasons, including the content or structure of the questionnaire or research process in addition to be that they are multi-tasking and participating in other activities at the same time.

Category: Participant Characteristics

Double Opt-In

Simple definition: A process to confirm the agreement of participants to opt in.

Full description: After signing up for a panel (opt-in) a participant will be sent a confirmation email. Double opt-in involves confirming an email address by responding to the confirmation email. The double opt-in process does not by itself provide sufficient protection against fraud.

Category: Panels and Samples

False Negatives

Simple definition: Incorrectly validated participants identified as of poor quality.

Full description: The proportion of genuine or good quality participants who are incorrectly categorised as poor quality or fraudulent participants.

Category: Participant Characteristics

False Positives

Simple definition: Incorrectly validated participants identified as of good quality.

Full description: The proportion of poor quality or fraudulent participants who are incorrectly categorised as valid.

Category: Participant Characteristics

Form Filling

Simple definition: A script or program to answer survey questions automatically.

Full description: An app, typically a browser plugin or server app, which automatically and randomly or algorithmically fills out survey questions quickly to save a fraudulent participant time completing a questionnaire.

Category: Behaviours and Biases

Fraudulent Participant

Simple definition: A participant who deliberately misrepresents their identity, profiling information or responses, including organisations that use bots to impersonate participants.

Full description: A participant who intentionally circumvents the research process, usually for monetary gain. This can occur in both qualitative and quantitative research. Most commonly, survey fraud occurs when participants attempt to collect survey rewards while a) not being qualified for the survey and/or b) making efforts to collect rewards while bypassing as much of the survey as possible. This includes but is not limited to: 1) not responding to questions honestly 2) taking surveys they are not qualified for 3) falsely posing as belonging to a particular demographic group or overclaiming at the screener 4) accessing surveys from countries not being targeted or in languages surveys are not being offered 5) intentionally taking surveys more than once from the same or multiple different accounts 6) the use of automation to generate closed-ended and open-ended survey responses 7) generation of false completes in participant system rather than actual survey data through manipulation of participant process flow (i.e. ghost complete) 8) using false ID to impersonate or create a profile

Category: Participant Characteristics

Generative AI

Simple definition: Refers to a branch of artificial intelligence that focuses on creating systems capable of producing new and original content.

Full description: This could include such things as images, music, or text. It involves training models to learn the underlying patterns and characteristics of a given dataset and then using that knowledge to generate new instances that resemble the original data. Generative AI models aim to mimic human creativity by generating novel outputs that have not been explicitly programmed, although human-like creativity is aspirational at this point.

Category: General Terms

Geo-Location Tracking

Simple definition: A process to detect participants' locations.

Full description: Identifying the physical location of participants (usually via IP address) to ensure participants are in the geographic locale they claim to be in. Many survey platforms can detect and reject participants that are not in the country or region they claim to be in. However, given the manipulation of IPs and other information, this may not be reliable.

Category: Tools and Methods

Ghost Complete

Simple definition: A complete survey response recorded in a participant system and not recorded as complete in the survey data.

Full description: The process by a person to create a complete in a participant system rather than survey system by manipulation survey links to generate a false complete. This is done solely for financial gain. This typically would not impact the survey data. There are solutions such as utilising server to server flows.

Category: Participant Characteristics

High velocity

Simple definition: Large volumes of responses to a research project in a short period

Full description: Term to describe when large number of responses are recorded over a short period of time in an atypical pattern. This could indicate fraudulent activity and checks should be made on the data looking for patterns e.g. by supplier source / sub-source, demographics, closed and open end responses

Category: Behaviours and Biases

HoneyPot (Bot Trap)

Simple definition: A deceptive technique/ trap used to counteract fraudulent use of information systems.

Full description: A HoneyPot, in the context of cybersecurity and internet security, is a deceptive technique used to detect, deflect, or counteract unauthorized use of information systems. It acts as a trap

designed to lure and deceive malicious actors, such as automated bots, spammers, or attackers attempting to exploit vulnerabilities in a system. A HoneyPot appears to be a legitimate part of a network or system, but it is actually isolated and closely monitored. It contains false or simulated data that would be of interest to attackers. For example, in research, it could be a "hidden question" that only a bot would see and answer. Activities within the HoneyPot are extensively monitored and logged to gather information about the tactics, techniques, and tools used by attackers. By attracting and engaging with malicious entities, a HoneyPot serves as an early warning system, alerting administrators to potential security threats and vulnerabilities. In some markets, a HoneyPot may refer to something that fraudulent participants seek out. In other words, in market research, the use of incentives creates a honeypot of money that fraudulent participants seek.

Category: Tools and Methods

In Survey Duplicates

Simple definition: Repeated data in a data set.

Full description: The same participants completing the same survey more than one time providing the same responses throughout a survey, or providing the same personal data when that information is collected. Intent may vary: is this someone purposely trying to take the survey multiple times or simply on multiple panels where they receive the same survey and have passed through any other duplication check e.g. using two devices through two different sources?

Category: Tools and Methods

Inattentive Participant

Simple definition: A participant who does not give an adequate level of thought to the responses they provide.

Full description: A participant who is distracted or disengaged and responds to questions without fully processing or understanding their content. There may be many reasons for the inattentiveness, for example, they might be watching TV while completing a survey and are not reading the questions carefully enough.

Category: Participant Characteristics

Incidence

Simple definition: A test to predict how many people may qualify for a survey.

Full description: Using known profiling data to create a "sample" to determine the number of people within a panel that will qualify for a survey. When a qualifying criteria is included that does not already exist as a data point in a participant's profile, known or estimated probabilities may be used to determine the incidence rate of the target participant within the panel. When probabilities are estimated, reaching full

target is not guaranteed.

Category: Panels and Samples

Intercept Sample/River Sample

Simple definition: Participants directed to surveys via website or other sources of advertising in real time.

Full description: Participants who participate in surveys via banners, video games and other ads. There is typically no opt-in process. River samples improve reach, but more care is required to ensure data quality.

Category: Panels and Samples

International Organization for Standardization (ISO)

Simple definition: The International Organization for Standardization (ISO) is an independent, non-governmental organisation that develops international standards

Full description: ISO creates international agreed standards for products, services and processes, which have been created by global experts. ISO standards provide organizations and businesses with international standards to achieve consistency in the development of products, services and processes.

Category: Quality Certifications, Accreditations and Associations

IP-Deduplication

Simple definition: A process to identify participants who are taking surveys from the same IP address.

Full description: Due to the easy and inexpensive access to VPNs that can spoof IP address, this is not truly reliable quality/fraud detection criteria. Anyone making more than a casual attempt at fraud would be likely to use multiple IP addresses. This also means that using IP address to verify location can be unreliable. It is also possible that more than one genuine participant may be on the same panel in the same household and would have the same IP address. Other measures should be factored in. It is also possible that it is very important to understand that IP address is considered personal information (PII) in many countries, and certainly within EEA markets and the UK where GDPR applies. Informed consent must be gained before IP address is captured for any purpose.

Category: Tools and Methods

ISO 20252:2019

Simple definition: An international standard setting out the vocabulary and services requirements for market, opinion and social research, including insights and data analytics

Full description: The ISO 20252:2019 standards establishes terms, definitions and service requirements for service providers conducting market, opinion and social research, including insights and data analytics (referred to as ""service providers""). Non-market research activities, such as direct marketing, are outside the scope of the standard.

Category: Quality Certifications, Accreditations and Associations

ISO/IEC 27001:2022

Simple definition: An international standard to manage information security

Full description: ISO/IEC 27001 is an information security management systems (ISMS). It defines requirements an ISMS must meet. The ISO/IEC 27001 is an international standard provides companies of any size and from all sectors of activity with standards for establishing, implementing, maintaining and continually improving an information security management system. Conformity with ISO/IEC 27001 means that an organization or business has put in place a system to manage risks related to the security of data owned or handled by the organization or business, and that the system meets the requirements of the international standard.

Category: Quality Certifications, Accreditations and Associations

Large Language Model

Simple definition: A specific type of GenAI model that is designed to understand and generate human language.

Full description: These models, such as ChatGPT and Bard, are trained on massive amounts of text data and can generate coherent and contextually relevant text responses based on plain language prompts or queries. They learn the statistical relationships between words and phrases in the training data and use this knowledge to generate human-like responses or create original written content. LLMs have a wide range of applications, including natural language understanding, translation, programming, chatbots, summarizations, and content generation. One of the risks to research is that these models can be used by fraudsters to program intelligent bots that provide good and valid answers to open-ended questions in surveys.

Category: General Terms

Length Of InterviewÂ (Lol)

Simple definition: The time taken to complete the survey.

Full description: Usually based on the median survey completion time. This is used to detect speeding which also can indicate problematic participants. The length of a survey can have a significant impact on engagement and quality which is why longer interviews are more often conducted by phone or face to face as these are more engaging data collection methods.

Category: Survey Design

Low Incidence Check

Simple definition: An attention checking survey question.

Full description: Disguised screener questions. For example, a very low incidence brand of toothpaste may be put prominently in a brand list to see if the incidence does not match. This can provide false positives e.g. halo effect and practitioners should always use multiple questions to identify the truly engaged from the unengaged.

Category: Tools and Methods

Machine Learning

Simple definition: Machine learning is the training of computer systems on a given data set to recognize specific patterns that exist in the data.

Full description: Machine learning is the training of computer systems on a given data set to recognize specific patterns that exist in the data. As new data is added, analysts can guide parameters to tune the model or allow the algorithm to train itself to make more accurate predictions. Machine learning is used to ingest large quantities of data to help humans identify patterns more quickly and accurately.

Category: General Terms

Matrix (AKA Grid) Questions

Simple definition: Closed-ended survey questions with a grid like column structure.

Full description: Closed-ended survey questions with a characteristic grid like column structure in which columns typically correspond to response-options. The number of questions in a grid and how many are presented on screen at once should be considered as an important factor in creating a good participant experience.

Category: Survey Design

Max Diff Questions

Simple definition: A question to evaluate pairs of items.

Full description: Results from a MaxDiff (Maximum Difference or Best/Worst Scaling) exercise can also be used to identify patterns of fraudulent responses. Further reading; <https://en.wikipedia.org/wiki/MaxDiff>.

Category: Tools and Methods

Mischievous Participant

Simple definition: A person who provides information that is intentionally false or misleading.

Full description: As opposed to fraud, mischievousness is less likely to be motivated by financial gain. It could be for personal or societal reasons as a protest. For example, 400,000 people reported their religion as Jedi in the 2001 British Census.

Category: Participant Characteristics

Non-Naivete

Simple definition: Being very familiar with survey content.

Full description: A response bias that results from familiarity with the survey intentions or content. This is most commonly observed among highly active participants who participate in many studies.

Category: Behaviours and Biases

Open Ends / Verbatim Comments

Simple definition: Type of question where participants are asked to answer in their own words.

Full description: Questions designed for participants to answer without a pre-populated answer set. Can be text or numeric. Generally designed to obtain unprompted responses and gain a deeper understanding of personal opinion. Open ended responses are used in a number of different types of data quality checks.

Category: General Terms

Open-Ended Response Validation

Simple definition: A process to check open-ended responses to determine the quality of participants.

Full description: Open-ended responses are an effective tool for measuring quality and detecting fraud. Controls should include the following:

- Gibberish and nonsense answers.
- Answers that are off topic and have nothing to do with the question.
- Checking for duplications between questions and participants.
- Checking the language. Is the answer written in the correct language?
- Has the answer been copy-pasted into the text field? (This can only be detected with third-party solutions)
- Was the answer generated by an AI model? (This can only be detected with third-party solutions)

It should be noted that some responses such as "gibberish" open ends cannot be taken as a Problematic Participant as some participants just do not like providing open ends and that should be accepted. Based on the type of open end response different actions are required including reviewing the open end responses.

Category: Tools and Methods

Overclaiming

Simple definition: Deliberately exaggerate to qualify for a survey.

Full description: For example, claiming to be a senior manager in an organisation when they are a junior manager, or claiming extremely high awareness of brands or products.

Category: Behaviours and Biases

Panel Management

Simple definition: Process by which panels and panellists are administered to ensure high quality and engaged participants.

Full description: Based on a panel's business demands, for example, in some markets, smoking is prevalent so profile data relating to this would need to be kept regularly updated compared to other markets where this may need less regular review. Panel management is used to monitor panel targeting/profiling to make sure it is consistent with qualifying or termination metrics. Each panel will have activity thresholds that allow for grouping panellists where different levels of engagement are required, which maximises the responses rates, making sure the panellists' experiences are good and relevant activities are delivered to them.

Category: Panels and Samples

Panel Opt-In

Simple definition: Individuals who have registered with panels as a participant. This involves providing an email address and potentially other personal data.

Full description: To “opt-in” to an online panel is to sign up for a panel as a participant. This involves providing an email address and potentially other personal data and information. It is during the opt-in stage that multiple checks are made to ensure the individual is a unique and valid potential member for an online panel. Validating a person's identity successfully now requires even greater time and cost due to the easy accessibility of multiple free email accounts.

Category: Panels and Samples

Panel Sample

Simple definition: Participants drawn from a panel to which they have opted in.

Full description: Participants recruited from a documented source who have provided profile data and appropriate information for validation of identity, given explicit consent to participate in research according to the terms and conditions of panel membership, and has not opted out. Although a national representative sample (or other types of representation) can be curated from within a panel using quotas, panels will still comprise those that are interested in taking part in surveys and the rewards offered for doing so.

Category: Panels and Samples

Participant (Survey) Experience

Simple definition: A participant's view on the quality of the interaction when completing a survey.

Full description: How a participant feels when participating in a survey. A participant can experience excitement, boredom, interest, frustration, and a wide range of other reactions that can affect their level of engagement, attention, and focus, influencing the quality of their responses. The experience is affected by all stages from the panel portal, survey invite, screening and profiling and the actual survey. In particular, a survey router process may involve being profiled and qualified more than once which can create frustration.

Category: Survey Design

Passive Data Collection

Simple definition: Collection of data without direct interaction with the participant.

Full description: The permission-based or ethical collection of data by researchers observing, measuring, recording, or appending a research subject's actions or behaviour for the purpose of research and without direct interaction with the research subject.

Category: General Terms

Personal Data

Simple definition: Personal information that is attributable to an individual.

Full description: Personal data (sometimes erroneously referred to as "personally identifiable information" or "PII" or "personal information") means any information relating to a natural living person that can be used to identify an individual, for example by reference to direct identifiers (such as a name, specific geographic location, telephone number, picture, sound, video recording or biometric data) or indirectly by reference to an individual's physical, physiological, mental, economic, cultural or social characteristics. The definition most widely used is that defined in the EU GDPR. "Sensitive personal information" is a sub-category of this and requires additional care during collection, transferring and processing. This is dependent on which country and laws apply to the data subject. For example, in the UK, the following are considered special category data: - race; - ethnic origin; - political opinions; - religious or philosophical beliefs; - trade union membership; - genetic data; - biometric data (where this is used for identification purposes); - health data; - sex life; or - sexual orientation

Category: General Terms

Positional Bias

Simple definition: Selecting a response option based on its position in a list.

Full description: When a participant preferentially selects a response based on its position. Most commonly, problematic participants select the first (or top) available response option. There are simple techniques to reduce the impact of this, notably randomising or rotating lists, or using alphabetical order where it makes sense.

Category: Behaviours and Biases

Pre Survey Duplicates

Simple definition: Survey entry attempts from the same individual.

Full description: The same participants directed to the same survey more than one time, as identified by checks on IP addresses and browser fingerprints; duplication is device based rather than participant based and assumes that the same participant is using the same device or providing the same personal data when that information is collected. Intent may vary: is this someone purposely trying to take the survey multiple times or simply on multiple panels where they receive the same survey? There can be false positives if two people share the same device e.g. within the same household.

Category: Tools and Methods

Pre-Screening Validation Questions

Simple definition: Initial questions in a data collection instrument used to establish suitability of

participants.

Full description: Best practice would be to disguise the topic of the research activity so as to minimize the risk of individuals making false claims to be included. A disguised screener, which includes high and low incidence categories, supported with data cleaning makes it easier to determine if an individual participant is overclicking (choosing everything, even if very low incidence) or underclicking (not choosing a reasonable number of categories given the incidence). In addition, systems like Captcha and ReCAPTCHA can be used to ensure a human participant is taking surveys.

Category: Tools and Methods

Pre-Survey Quality Validation

Simple definition: A type of behavioural activity to verify the quality of participants before surveys are completed.

Full description: The process of identifying and removing low quality participants before they enter a survey. For example, digital fingerprinting, validation against third party financial fraud databases, IP address checks, exclusion lists.

Category: Tools and Methods

Privacy Laws

Simple definition: Laws that regulate the processing of personal data in any way.

Full description: The body of laws that deals with the regulating, storing, and using personal information which can be collected by governments, public or private organizations, or other individuals. CPRA and GDPR are two examples of a privacy law. Laws vary greatly around the world and even within countries like the USA. Care should be taken to ensure compliance in the country the work is being conducted from as well as in the country in which the individual resides.

Category: General Terms

Probabilistic panel

Simple definition: A panel which contains participants and utilizes a randomized selection process where every person in the target population theoretically has an equal chance to be selected as a participant

Full description: Panels which are created by choosing participants to recruit (versus opt-in). Probabilistic panels are recruited using random probability sampling, either via unclustered address-based or computer assisted telephone sampling. Individuals and/ or households need to be invited to join, versus simply volunteering. The objective of probabilistic recruitment is to ensure that every individual/household in a population of interest (such as a country) has an equal chance of being selected

to join a panel to support the statistical reliability of data collected.

Category: Panels and Samples

Problematic Participant

Simple definition: An umbrella term for all forms of bad quality participants.

Full description: This includes those who use certain types of automation such as translation apps. Not all forms of problematic participants should be categorized as fraud. For example, inattention may be considered as problematic participant behaviour rather than fraud.

Category: Participant Characteristics

Professional Survey Taker (Professional Participant)

Simple definition: A person who participates in surveys (or research generally) as a hobby or source of income.

Full description: There are two main types of professional survey taker:

- Bad intent: One simply participating in surveys at a high rate but with no regard for validity or quality of responses. They would usually be doing this to maximise the financial return.
- Good intent: An active participant that provides valid and quality responses and enjoys taking part in surveys and giving their opinions. It is important to not exclude those with good intentions purely based on frequency. Other quality checks should be factored in.

A consideration relating to all professional survey takers is "Non-Naivete" which is a response bias that results from familiarity with the survey intentions or content. This is most commonly observed among highly active participants who participate in many studies. This may result in participants that can predict what answers may qualify them for the study or may generate higher rewards.

Category: Participant Characteristics

Qualitative Research

Simple definition: Collection and analysis of open ended unstructured data used to develop insights.

Full description: An unstructured research approach with a small number of carefully selected individuals used to produce non-quantifiable insights into behaviour, motivations and attitudes. Generally conducted by telephone or face to face either online or in person but in some cases using live online text/image based discussion platforms. Qualitative generally involves using small sample sizes and a longer length of interview, typically between 30 and 60 minutes. Qualitative research shares a number of data quality concerns with quantitative research.

Category: General Terms

Quantitative Research

Simple definition: Collection and analysis of structured data.

Full description: Research centred around the numerical variations held within the dataset and based on statistical outcomes usually in the form of percentages. Surveys are made up of single/multiple choice questions, a small number of open ends, scales, ranking etc. Survey methodology that employs closed-ended response options that are easily numericized. Quantitative research is particularly vulnerable to fraud due to the ease with which closed-ended responses can be made by inattentive participants and automated form fillers, and the relative difficulty in assessing the quality of closed-ended responses. Attention to the speed of completion, answer patterns and quality of verbatim comments can be used to assess whether fraud is taking place.

Category: General Terms

Question Design

Simple definition: How a question is designed to elicit the appropriate type of response from participants.

Full description: Elements of question design that affect how and in what manner participants understand and respond. Various aspects of question design can affect participant experience, including whether it is easy to understand, includes an answer option for all possibilities, is skipped for individuals for whom it should be skipped, is leading and or double-barrelled, is cognitively hard to complete.

Category: Survey Design

Random Response Profile

Simple definition: Selecting random and unconsidered responses.

Full description: When a participant randomly selects from among a question's available response options. Note that this is different from straight lining or acquiescence bias, which are not random, and are more easily detectable. Form-fillers at times employ a "random or semi-random response strategy."

Category: Behaviours and Biases

Re-CAPTCHA / CAPTCHA (bot checks)

Simple definition: Tools designed to protect online ecosystems from automated abuse, ensuring transactions are with humans rather than bots.

Full description: CAPTCHA stands for "Completely Automated Public Turing test to tell Computers and Humans Apart." It is a challenge-response test used in computing to determine whether the user is human or automated software (often referred to as bots). The primary purpose of CAPTCHA is to prevent automated bots from performing actions that should be reserved for humans, such as submitting forms, creating accounts, or conducting transactions. CAPTCHA typically presents a task or puzzle that is easy for a human to solve but difficult for automated scripts or bots, such as Image Recognition (e.g., selecting all images with traffic lights), Text-Based Challenges (e.g. type in distorted text displayed in an image), Mathematical Problems (e.g., solving a basic addition or subtraction equation). reCAPTCHA protects online systems and applications by detecting and protecting accounts from hacks, blocking credential stuffing attacks, and creating accounts in bulk, all without user friction. It uses an invisible score-based detection mechanism to differentiate legitimate users from bots and other malicious attacks. In summary, CAPTCHA and ReCAPTCHA are tools designed to protect websites and online services from automated abuse by ensuring that interactions are performed by human users rather than bots.

Category: Tools and Methods

Red Herring/Explicit Trap Question

Simple definition: An attention checking survey question.

Full description: A question in a survey designed to check whether participants are paying attention. A common example instructs participants to select a "strongly disagree" response option. Sometimes known as an INSTRUCTIONAL MANIPULATION CHECK (IMC). These can generate false positives and typically not used as a stand alone check.

Category: Tools and Methods

Representativeness:

Simple definition: Degree to which a sample reflects the target population being studied.

Full description: As with "sample" - ensuring that a representative snapshot of the larger target population is obtained in order to guarantee a like for like measure of opinion. Quota sampling is required as a means of obtaining an illustrative data set that represents the larger potential target market. A good example is "Nat Rep" (Nationally Representative) where the specific project target requires full coverage of all groups that feature within a country's census data for example - gender, age, region etc.

Category: General Terms

Response Patterns For Repeated Question Sets

Simple definition: Systematic responses to survey question banks / grids.

Full description: When a participant answers a bank or grid of survey questions in a detectable pattern

that indicates a lack of attention or deliberate fraud. This includes but is not limited to straight lining/flat lining as a participant may select response options in a non-linear pattern that is harder to detect. This is particularly hard to detect in data sets where the order of questioning is randomised. Other factors such as time taken across the particular question set can be utilised.

Category: Behaviours and Biases

Router

Simple definition: Technology that redirects participants to specific surveys.

Full description: An online software application that screens incoming research participants and then uses those results to assign participants to one of multiple available research projects. A router can also offer participants additional screeners and surveys after screener qualification failure or survey completion. As many panel organisations work with partners or partner networks, technology has been developed to automate the process of passing a panel member from their source panel, via the panel managing the project and ultimately to surveys. During this routing process, participants may have to answer multiple, and sometime repetitious profiling questions that can create a negative experience. Panel organisations and the research sector are working together to create guidelines on how to align profiling data to reduce repetition and optimise this process.

Category: Panels and Samples

Sample (Survey Sample)

Simple definition: A subset of the target population from which quantitative data are collected to represent a larger population.

Full description: It is a subset containing the characteristics of the larger target population but still sizeable enough to allow for robust analysis. The subset of the population or universe of interest which is interviewed. Typically, the sample is pulled using random methods, so that everyone has an equal chance of inclusion. However quotas are set in order to guarantee representativity in terms of gender, age, geolocation etc in addition to the specific characteristics of the project design, for example - primary grocery shopper, automotive purchase decision maker, C-suite etc.

Category: Panels and Samples

Satisficing

Simple definition: When research participants use minimum effort required to answer surveys

Full description: Satisficing in the context of research refers to the tendency of participants to provide satisfactory or 'good enough' answers rather than carefully considered or optimal responses. This behaviour occurs when participants choose to expend minimal effort or cognitive resources to complete a survey or questionnaire due to lack of motivation or because the survey task is too difficult.

Category: Behaviours and Biases

Speeding/Racing

Simple definition: Completing a questionnaire extremely quickly.

Full description: Extremely fast survey completion times. Thresholds for what is considered "every fast" can be based on specific minimum completion times (e.g. faster than 1 minute in a study with a median completion time of ten minutes). Many studies show that some valid participants are capable of very fast completion times, and care must be taken to avoid false positives. Thus, speeding is often used as a flag, rather than as a sole rejection criterion unless the completion time is considered impossibly fast. Speed can be measured at a question level and multiple question timings can be combined to create a more accurate picture of the behaviour of a participant. Fraudulent Participants may complete the bulk of questions extremely fast but then pause on the last question to deliberately increase the overall completion time.

Category: Behaviours and Biases

Staged fielding

Simple definition: Staged fielding is a strategic approach to collect data across multiple phases rather than one.

Full description: Staged fielding refers to a strategic approach where data collection for a study is conducted in multiple phases or stages rather than all at once. This method allows researchers to refine and improve their survey instruments, methodologies, sampling strategies or fraud-mitigation strategies based on insights gained from earlier stages of data collection. Data collection can be divided into sequential stages, with each stage focusing on a specific aspect of the research process or used to gather feedback and insights from each stage of fielding to make adjustments and improvements to subsequent stages. Staged fielding allows researchers to mitigate risks associated with large-scale data collection by identifying and addressing problems early in the research process. This can include issues related to survey design, respondent engagement, or data quality.

Category: Tools and Methods

Straight Lining/Flat Lining

Simple definition: Providing the same answer to the majority of survey grid questions.

Full description: This is a type of Response Pattern. While this behaviour can be easily detected in survey data, using this as a reason for rejection needs to be carefully considered. A participant may do this if, for example, a question is not relevant to them, or if they have become disengaged for some reason. They may actually be answering honestly but just happen to have the same answer each time.

The specific threshold can depend on the nature of the study and should be set accordingly based on the number of survey grid questions, and number of items in grids. The term "straight lining check" should be discontinued and replaced with "response pattern check".

Category: Behaviours and Biases

Survey Design

Simple definition: The questionnaire structure and methods used to gather data.

Full description: Survey design is the structure of the questionnaire and any specific techniques (methods) leveraged to capture data.

Category: Survey Design

Synthetic Data

Simple definition: Information that is artificially created to augment or replace real data.

Full description: Synthetic data is artificial data that is generated from a model, which could use original data, that is trained to reproduce the characteristics and structure of realistic data. Synthetic data can be created via three methods, including via: - Large Language Models, - Ascription which is copying interview data onto another interview and comparing similarities via the measurement of distances, and - Machine Learning which trains a model using original data. The objective is to ensure synthetic data and original data deliver very similar results when undergoing the same statistical analysis. That being said, this does require careful management and monitoring, and it can be the case that the results are different from expectations. The degree to which synthetic data is an accurate proxy for the original data is a measure of the utility of the method and the model. Synthetic data is classified by the types of original data used to create the data e.g. real datasets, analytical data based upon models, etc. Generative AI tools can be used to create the synthetic data. Note: Synthetic Data is quickly evolving. As such, this definition will continue to evolve. It is recommended that practitioners do their due diligence with regards to its use.

Category: Participant Characteristics

Synthetic Participant

Simple definition: Artificially created participants to augment or replace real participants.

Full description: Synthetic participants are artificially created data records designed to simulate real participants using techniques such as generative AI. They can supplement an existing human sample to fill quotas or increase sample size, or they can simulate an entire survey independently. Note: The use of Synthetic Participants is quickly evolving. As such, this definition will continue to evolve. It is recommended that practitioners do their due diligence with regards to the usage of such data.

Category: Participant Characteristics

System Virtual Machine

Simple definition: A software emulation of a physical computer system that runs an operating system (OS) and applications.

Full description: A System Virtual Machine (VM) is a software emulation of a physical computer system that runs an operating system (OS) and applications. Unlike application virtual machines, which are designed to execute specific software programs, a system virtual machine provides a complete virtual environment that mimics the hardware and software configuration of a physical machine. Fraudsters use VMs to mask their true identity and location that allow them to mimic a different device than what they're actually using and pass security checks during digital fingerprinting.

Category: Tools and Methods

Underclicking

Simple definition: Deliberately choosing fewer items in a list to prevent repetition.

Full description: For example, choosing 2-3 brands in a list where many would be well known. This would be to avoid repeated loop questions about each brand.

Category: Behaviours and Biases

Valid Participant

Simple definition: A person who is engaged, honest, and meets research participation requirements.

Full description: This is the ideal participant but great care should be taken to identify a valid participant as, for example, what may look like an unengaged participant may be a result of research being poorly designed or worded, or a survey being too long. Multiple data points and factors should be taken into account when determining whether a person is a valid participant or not.

Category: Participant Characteristics

Validation Of Fraud-Detection Solutions

Simple definition: A process to determine the effectiveness of solutions used to identify inappropriate activity.

Full description: Empirical assessment of the effectiveness of solutions at identifying and blocking

problematic participants. Such assessment should include a description of methods and outcomes such as false positives and false negative rates.

Category: General Terms

Verification

Simple definition: The process of validating a person's identity.

Full description: Establishing a participant's identity or background via personal information or qualifying questions is crucial for online panels and other participant sources to combat fraudulent participation. Some panels, especially B2B panels, incorporate or plan to incorporate official ID verification. The verification level varies based on its placement in the survey process and who controls it. Proprietary research panels may employ more personal data for verification compared to other sources. Sample exchanges rely more on technical factors like browser or device information, although these are not foolproof measures and rather serve as mitigating strategies.

Category: Tools and Methods