

The Measurement of Sensitivity

Inaugural-Dissertation

Zur Erlangung des akademischen Grades einer Doktorin

der Wirtschafts- und Sozialwissenschaften

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Franziska Dunkelmann, M.A.

aus Wismar

Kiel, 2024

Erstbegutachtung: Prof. Dr. Peter Graeff

Zweitbegutachtung: Prof. Dr. Claudia Buengeler

Tag der mündlichen Prüfung: 16. Januar 2025

CONTENTS

1	Synopsis	1
1.1	The dissertation's general topic and it's relevance	1
1.2	Extended Abstracts	10
1.2.1	First Paper: Corruption and COVID-19. Possible similarities between two survey topics	10
1.2.2	Second Paper: Vignette studies	17
1.2.3	Third Paper: Disentangling Differences in Perceived Sensitivity and Misreporting Behaviour: A Factorial Survey	21
1.3	Relationship of the Papers.....	25
1.4	How the dissertation fills previous research gaps.....	26
1.5	References	27
2	First Paper.....	37
2.1	English Transcript.....	37
2.2	German Original.....	55
3	Second Paper	75
3.1	English Transcript.....	75
3.2	German Original.....	95
4	Third Paper	118
5	Declaration on independent work	167
6	Declaration about the Co-Authorship and the DFG-Guidelines.....	168

LIST OF TABLES

Table 1: Components of the response process	3
Table 2: Examples of population surveys using potentially sensitive items	6
Table 3: List Experiment from the National Race and Politics Survey 1991	9
Table 4: Examples of empirical publications with list experiments since 2018	12
Table 5: Example Vignette	17
Table 6: Example of a vignette presented to respondents, including questions to assess	23

1 SYNOPSIS

1.1 The dissertation's general topic and it's relevance

"Despite 70 years of methodological innovation, it remains difficult for researchers to decide whether sensitivity bias is likely to be a problem and if it is, how best to address it." (Blair et al., 2020: 1297)

In this quote Blair et al. refer to the main subject of this dissertation - sensitivity - by addressing the problematic nature of the concept. In order to understand sensitivity, a definition will be given. Subsequently, the concept of sensitivity bias will be addressed as well.

In the field of sensitivity research, the distinction proposed by Tourangeau and Yan (2007) has emerged as the most widely used definition of sensitivity. The two authors differentiate between three dimensions of sensitivity: (1) intrusiveness, (2) threat of disclosure and (3) social desirability (Tourangeau & Yan, 2007, p. 860; Tourangeau, 2018, p. 503; Yan, 2021, p. 111).¹

The first dimension includes topics that are regarded as social taboos and that are avoided in regular conversations. Respondents may perceive questions on such topics as an invasion of privacy and as inherently offensive because they delve into social taboos (Tourangeau & Yan, 2007, p. 860). It should be noted that the questions themselves already appear inappropriate and to a certain extent a *"question threat"* (Tourangeau et al., 2000, p. 258) emerges. The potential answer of respondents is irrelevant for this dimension of sensitivity and a sensitive question in this regard is associated with the risk of alienating all respondents. As a result, respondents could react to these questions with a general rejection. These questions can result in a refusal to answer or, in the worst case, even a dropout of the survey (Tourangeau & Yan, 2007, p. 860).

¹ This categorisation was already used by Tourangeau, Rips and Rasinski (2000, Chapt. 9). However, the publication by Tourangeau and Yan (2007) was used much more frequently in the research community and the distinction is often described as an invention from 2007.

Sensitivity can also arise in relation to the response options associated with a question. The second dimension of sensitivity is referred to as *"threat of disclosure"* (Tourangeau & Yan, 2007, p. 860; Tourangeau, 2018, p. 503; Tourangeau et al., 2000, p. 258f). According to this dimension, a question is sensitive if respondents fear that their answer will be disclosed to people not directly involved with the survey or other third parties (Tourangeau et al., 2000, p. 258; Tourangeau, 2018, p. 503). More specifically, sensitivity refers to a certain answer option where the respondent's truthful answer would reveal a confession of a certain behaviour. If this behaviour or answer becomes known to third parties, this could lead to negative consequences for the respondent (Tourangeau & Yan, 2007, p. 860). For this reason, there is a risk or threat of disclosure (Tourangeau et al., 2000, p. 258). Tourangeau and Yan (2007, p. 860) describe this dimension as an important factor in the misreporting of illegal or socially undesirable behaviour.

Tourangeau and Yan (2007, p. 860) also devote a separate, third dimension of sensitivity to the concept of social desirability. This third dimension focuses on applicable and perceived social norms. Social desirability can be understood as the sensitivity of the response options (Tourangeau, 2018, p. 503). Respondents compare their potential answer with the applicable norms and *"sensitive questions ask, in effect, whether we have violated such norms"* (Tourangeau et al., 2000, p. 257). Similar to dimension two (threat of disclosure), the effect of social desirability differs depending on the respondent's true answer. For respondents whose true answer is in accordance with social norms, there is simply no incentive to withhold this conformity. However, respondents whose behaviour, opinion or attitude deviates from current social norms - in other words: whose answer is socially undesirable - may have an incentive to edit their answer. These respondents could ponder whether to give their true and socially undesirable answer to a question, or to give a socially desirable answer instead. The consideration process described above assumes that there are clear social norms regarding a survey item and that respondents want to present themselves in a favourable light (Tourangeau et al., 2000, p. 257; Tourangeau & Yan, 2007, p. 860).

Even one of the mentioned aspects leads to the characterisation of sensitivity. Nevertheless, two or even all dimensions can occur simultaneously (Yan, 2021, p. 111). This can be illustrated by items covering respondents' sexuality. Questions on this kind of topic are taboo in many countries. At the same time, questions about sexuality are characterised by certain social

norms, such as fidelity to one's partner or the use of condoms to prevent sexually transmitted diseases. The aspect of socially approved fidelity in partnerships or the use of condoms can also cause a threat of disclosure if an answer in a survey admits a deviation from these norms. Questions about the sexual behaviour of respondents can be influenced by additional aspects that can be assigned to Tourangeau and Yan's dimensions. Even this small insight illustrates that the dimensions of sensitivity can not to be regarded as distinct categories.

Recalling the quote from the beginning of this chapter, besides sensitivity it also mentioned a *sensitivity bias*. This term already indicates that the sensitivity of a question or the answer options could lead to biased answers from respondents. This thought can be illustrated by taking a closer look at the response process. Upon reviewing the research literature, the response process can ideally be described as a stepwise procedure that respondents undergo more or less straightforward. More precisely, Tourangeau, Rips, and Rasinski (2000, p. 7-16) detailed the process of answering a question through four key components:

1. Respondents need to carefully interpret the meaning of each question.
2. Respondents have to retrieve all relevant information from their memories.
3. Respondents should thoughtfully integrate that information into concise assessments.
4. Respondents articulate those assessments in a manner that conveys meaning with clarity and precision.

This four-step framework describes a perfect response process with up to 13 individual steps (see table 1):

Table 1: Components of the response process

Component	Specific Processes
Comprehension	Attend to questions and instructions
	Represent logical form of question
	Identify question focus (information sought)
	Link key terms to relevant concepts
Retrieval	Generate retrieval strategy and cues
	Retrieve specific, generic memories
	Fill in missing details
Judgment	Assess completeness and relevance of memories
	Draw inferences based on accessibility
	Integrate material retrieved
	Make estimate based on partial retrieval
Response	Map judgement onto response category
	Edit response

Source: Tourangeau et al. 2000, p. 8

However, respondents can also - for various reasons - deviate from this stepwise procedure by completing steps incompletely or skipping entire steps (Krosnick, 1991, p. 215; Krosnick, 1999, p. 547f).

For instance, Tourangeau et al. 2000 (p. 8) identify the respondent's motivation to give an accurate response or the speed with which an answer must be produced as factors influencing the respondent's response process. In addition, sensitivity can also influence the response process. Sensitivity can manifest itself in various ways in the response behaviour of respondents.

In particular, respondents can react to a particular (sensitive) survey item in various ways (Tourangeau & Yan, 2007, p. 877; Tourangeau et al., 2000, p. 261-265)²:

- I. Respondents can refuse to answer if a sensitive question or a question with a sensitive answer option is asked (item non-response).
- II. Respondents may give a biased answer if a sensitive question or a question with a sensitive answer option is asked (misreporting).
- III. Respondents can answer truthfully even though a sensitive question or a question with a sensitive answer option is asked.

The possibility of misreporting is particularly relevant for research because this response bias (in contrast to the item non-response) is not visible. Biased responses can in general arise - also with reference to the work of Krosnick (1991, p. 215-220) or the response process according to Tourangeau et al. 2000 (p. 264) - as a result of a conscious or unconscious process. It is assumed that the sensitivity of a question or answer can lead to deliberate misreporting and that, in the case of sensitivity, respondents often rationally decide to give a biased answer (Celhay et al., 2022, p. 5; Fox J. A., 2016, p. 4; Tourangeau et al., 2000, p. 264f; Tourangeau & Yan, 2007, p. 876). At the same time, there is no automatism. In many cases, it is uncertain to what extent respondents will give a biased answer or answer truthfully despite sensitivity when confronted with a direct question.

² The sensitivity of a topic can also lead to a complete denial to participate in a survey (unit non-response) if, for instance, the topic is already mentioned in the invitation to a survey. This case is deliberately excluded here and the general willingness to participate in a survey is assumed.

Given the possible inaccuracy of self-reports on sensitive topics, the question may arise why they nevertheless play a central role in social science research. This will be recalled using some examples.³ Table 2 shows selected items from five population surveys. Looking at these examples in detail underlines some of the difficulties of alternative data on the phenomena.

Aggregate data from other sources can be found for various topics listed in the table. This applies, for instance, to questions about criminal offences (e.g., driving under the influence of alcohol or tax evasion from the ALLBUS, see table 2) or the prevalence of diseases (e.g., items from the ESS, see table 2). The use of aggregate data can generally generate worthwhile scientific progress. Nevertheless, the investigation of such phenomena at an individual level, for instance in the case of a research interest that investigates motivations for behaviour, also represents an important strand of research. As individual data is often only available to a very limited extent or not at all for data protection reasons, this research at the individual level needs surveys as the basis of research.

Furthermore, the available aggregate data is often known as “bright field” data. This data only contains cases that are known to an institution or authority. In the case of criminal offences these are only the reported or convicted offenders, whereas the undetected offenders belong to the so-called “dark field” (Heimann, 2022, p. 22ff). In the case of (mental) illnesses, data from health insurance companies are available in some cases. However, these only contain patients who are treated for an illness. In addition, data from health insurance companies can only be meaningful in some countries (with a comprehensive system of health insurances). This example also illustrates the relevance of surveys. On the one hand, it is possible for undetected offenders or untreated patients to self-report their offence or illness. On the other hand, for instance in countries without comprehensive health insurance, a survey is even more necessary in order to determine the spread of diseases or any changes in this regard.

³ This list is non-exhaustive and is not intended to fully undermine the usefulness of data besides surveys. The examples are merely intended to illustrate the need for surveys and self-reports.

Table 2: Examples of population surveys using potentially sensitive items

German General Social Survey (ALLBUS, 2021) Source: GESIS Leibniz-Institut für Sozialwissenschaften, 2023 As you know, many people now and then commit minor offences. Below are four such minor offences. Please indicate for each offence how often you have already done this in your life. • Used public transport without having a valid ticket. • Drove a car with more than the permitted level of blood alcohol. • Took goods from a department store without paying. • Made false statements on your income tax return in order to pay less tax. ○ Never ○ 1 time ○ 2 to 5 times ○ 6 to 10 times ○ 11 to 20 times ○ More the 20 times
European Social Survey (ESS, round 10, 2020/21) Source: European Social Survey, 2020 Are you hampered in your daily activities in any way by any longstanding illness, or disability, infirmity or mental health problem? IF YES, is that a lot or to some extent? ○ Yes, a lot ○ Yes, to some extent ○ No
General Social Survey (GSS, 2022) Source: National Opinion Research Center, 2023 Have you ever, even once, taken any drugs by injection with a needle (like heroin, cocaine, amphetamines, or steroids)? DO NOT include anything you took under a doctor's orders. ○ Yes ○ No
World Values Survey (WVS, wave 7, 2017-2021) Source: World Values Survey Association, 2022 In the last 12 months, how often have you or your family... • Gone without enough food to eat. • Felt unsafe from crime in your home. • Gone without medicine or medical treatment that you needed. • Gone without a cash income. • Gone without a safe shelter over your head. ○ Often ○ Sometimes ○ Rarely ○ Never
The International Social Survey Programme (ISSP, 2018) Source: ISSP Research Group, 2020 What is your attitude towards people from the following religious groups? • Christians • Muslims • Hindus • Buddhists • Jews ○ Very positive ○ Rather positive ○ Neither positive nor negative ○ Rather negative ○ Very negative

Official statistics or comparable data are still not available for all types of behaviour. This applies, for instance, to the item on drug use (from the GSS, see table 2). In most countries, dealing and/or drug possession is prosecuted, so that bright field data on these criminal offences are also documented. In contrast, the consumption of drugs can be measured using surveys. Furthermore, surveys on experiences with drug use could be used to systematically and more directly indicate changes, such as an increase in the use of a particular substance.

Another aspect is illustrated by the items from the World Values Survey (see table 2). The items listed cover a range of circumstances that are strongly related to the respondent's private life or to incidents in the respondent's home. For this context, alternative data sources are hardly conceivable or can be realised for rather small(er) samples, for instance with the help of other social science methods like observations. In addition, data for one example - the item on the feeling of insecurity - cannot be collected even in the context of observations.

This applies in particular to every phenomenon regarding perceptions, opinions and attitudes of respondents. This field of research relies on surveys to obtain data because there cannot be aggregated data or observational data for invisible perceptions, opinions or attitudes (e.g., see item from the ISSP in table 2).

Summarising the examples, it becomes clear that although surveys and self-reporting can be problematic, they are still necessary for social science research. For this reason, specific survey methods were developed for sensitive topics.⁴ These are intended to enable self-reporting by respondents despite sensitivity and have already been applied to various subject areas. Two examples that have been used more frequently are the Randomised Response Technique (RRT; by Warner, 1965) and the Item Count Technique (ICT, also known as the List Experiment or Unmatched Count Technique (UCT); by Miller, 1984).

In RRT, which occurs in different variants, a random component is integrated into the survey process (Warner, 1965, p. 63; Wolter & Preisendörfer, 2013, p. 322). These methods can be referred to as either forced response or forced questions variants (John et al., 2018, p. 102).

⁴ It is important to note that a randomly distributed bias according to sensitivity would not be problematic. In such instances, the use of direct questioning would remain a suitable option. The following methods were developed with the objective of reducing systematic sensitivity bias.

For instance, respondents are requested to toss a coin or roll a dice and then refrain from disclosing the result. The next step depends on the variant of RRT:

- I. Forced question variant (e.g., Greenberg et al., 1969): Respondents are asked two questions. One is the sensitive item of interest and the other is non-sensitive. The result of the coin flip or dice roll determines whether the respondent answers the sensitive question or the alternative item.
- II. Forced response variant (e.g., Warner, 1965): All respondents are asked the sensitive question. The result of the coin flip or dice toss determines whether the respondents answer the sensitive question truthfully or are instructed to give a specific answer.

After the survey there is an individual answer for each respondent, but it is not possible to assign which question was answered (I) or whether the answer is truthful or not (II). However, the prevalence of the sensitive item can be estimated for all respondents using the probabilities of the outcomes of the random mechanism (Fox J. A., 2016, p. 13f). This increase in anonymity is accompanied by a higher standard error in the estimate (Droitcour, et al., 1991, p. 187; Fox J. A., 2016, p. 73). Nevertheless, the procedure is intended to mitigate the effects of sensitivity or ideally even avoid them altogether, because unpleasant answers can be given more anonymously. This increase in anonymity should increase the response probability or the willingness to answer truthfully with regard to the sensitive item (Lensvelt-Mulders, et al., 2005, p. 321; Warner, 1965, p. 64).

With ICT, respondents are also not asked any direct questions if sensitivity is assumed. Instead, respondents are shown a list of questions with the same answer format. Respondents are explicitly asked not to indicate which statement they agree or disagree with. Instead, respondents indicate for how many items this is the case. In this experimental design, the respondents are also divided into two groups. The control group receives a list without the sensitive item (referred to as the *short list*) and the treatment group receives the same list plus the sensitive item (referred to as the *long list*). The first list experiment carried out in a population survey and with two treatment groups illustrates this procedure:

Table 3: List Experiment from the National Race and Politics Survey 1991

List for the control group	List for the treatment group 1	List for the treatment group 2
Now I'm going to read you three things that sometimes make people angry or upset. After I read all three, just tell me HOW MANY of them upset you. (I don't want to know which ones, just how many.) (1) the federal government increasing the tax on gasoline; (2) professional athletes getting million-dollar-plus salaries; (3) large corporations polluting the environment. How many, if any, of these things upset you?	Now I'm going to read you four things that sometimes make people angry or upset. After I read all four, just tell me HOW MANY of them upset you. (I don't want to know which ones, just how many.) (1) the federal government increasing the tax on gasoline; (2) professional athletes getting million-dollar-plus salaries; (3) large corporations polluting the environment; (4) a black family moving in next door to you. How many, if any, of these things upset you?	Now I'm going to read you four things that sometimes make people angry or upset. After I read all four, just tell me HOW MANY of them upset you. (I don't want to know which ones, just how many.) (1) the federal government increasing the tax on gasoline; (2) professional athletes getting million-dollar-plus salaries; (3) large corporations polluting the environment; (4) black leaders asking the government for affirmative action. How many, if any, of these things upset you?
Note: sensitive item highlighted Source: Sniderman et al., 1992		

As this example shows, a list experiment is a design in which respondents answer a list of items simultaneously. In the classic design, the items on the list and therefore also the sensitive item are formulated in such a way that respondents can easily count how many items they agree with or how many they disagree with⁵. Nowadays, the lists often contain at least one item that the majority of respondents agree with and one that the majority of respondents disagree with. Alternatively, the combination of two negatively correlated items has proven to be useful (Glynn, 2013, p. 163). The objective of this procedure is to prevent floor or ceiling effects, which are defined as the agreement with all or none of the statements. As long as no floor or ceiling effect occurs, the respondent's answer to the sensitive item remains hidden in this design. The respondents do not have to express a clear opinion on a sensitive topic or disclose a sensitive behaviour. However, the prevalence of the sensitive item of interest can

⁵ One limitation of this design was the restriction to the specific wording of the statements. In a list experiment, the wording of the items is initially limited to the nominal scale level (usually disagreement or agreement). An extension of the design, the so-called Item Sum Technique (Trappmann et al., 2014) also includes questions that are asked at the absolute scale level.

be estimated by a mean value comparison between the treatment and control group.⁶ This increase in anonymity also induces uncertainty and the estimate reflects this with a higher standard error than an estimate based on a direct question (Droitcour et al., 1991, p. 189; Imai, 2011, p. 407).

The basic assumption that unites the survey methods is that respondents will provide information on sensitive topics (behaviour and attitudes) if the response is not immediately obvious even to the interviewer or researcher. In summary, these two methods have been applied to a range of sensitive topics (meta-analyses e.g., Blair et al., 2020; Krumpal, 2013; Lensvelt-Mulders et al., 2005). However, both methods are clearly more costly and time-consuming to implement, which is why integration has so far been limited to some subject areas and research projects.

To summarise, there are considerable gaps in the research on sensitivity and the effects of sensitivity. In many cases, sensitive survey topics are captured in direct questions, but it remains unclear how reliable the data obtained in this way is.

1.2 Extended Abstracts

1.2.1 First Paper: Corruption and COVID-19. Possible similarities between two survey topics

The first article of my dissertation was written in 2020 and pursues a theoretical approach. The article systematically addresses the question of whether items on corruption and anti-corruption regulations can be categorised as sensitive. A structured comparison of items surveying compliance in relation to COVID-19 demonstrates that a survey paradigm can be employed to collect self-reports on the topic of corruption despite the assumed sensitivity of the issue.

⁶ Based on this, methods for multivariate analyses were also developed (e.g., see Imai, 2011).

As already described in the previous chapter, list experiments or the ICT are a specific survey method that was developed to collect self-reports on sensitive topics. This method has already been used many times, as Blair, Coppock and Moor show in their meta-study from 2020. The authors themselves emphasise that they could not provide a complete collection of all the list experiments that have ever been conducted. Nevertheless, the authors were able to identify a total of 487 distinct experiments in 154 different publications by the end of 2017 (Blair et al., 2020, p. 1306).

In a further meta-analysis, Ehler, Wolter, and Junkermann (2021) also examined the performance of list experiments in comparison to direct questioning on sensitive topics. The data basis included list experiments conducted up to the beginning of 2019. The authors argue in favour of list experiments compared to direct questions (Ehler et al., 2021, p. 23).

A self-generated list of exemplary studies published in the last five years emphasises that list experiments continue to be conducted and have thus become an even more established survey paradigm (see Table 4):

Table 4: Examples of empirical publications with list experiments since 2018

Topic	Wording of the sensitive item	Source	Sample
Support for an ongoing war	How many of the following things do you personally support? [...] Actions of the Russian armed forces in Ukraine.	Chapkovski & Schaub, 2022	Russia (N = 3000)
Considering Chinese immigrants as a threat	Chinese immigrants damage the public health in the U.S.	Xu & Liu, 2021	United States (N = 1200)
Consumption of bear bile products	Bought or used bear bile products.	Hinsley et al., 2021	China (N = 3646)
Experience of sexual violence during civil war (indirectly including experience of homosexuality)	Now we would like to ask you some more questions about what happened during the war. Please refer to the following list and tell me how many of these experiences happened to you during the war. [...] I was personally sexually assaulted.	Traunmüller et al., 2019	Sri Lanka (N = 1800)
Tax evasion	I have reported an amount lower than my actual income in my tax report.	Iraman et al., 2022	Jakarta, Indonesia (N = 879)
Rejection of a presidential candidate belonging to minorities	How many of the following characteristics would make you more likely to vote against a candidate for President? [...] - A candidate who is African American. - A candidate who is gay or homosexual. - A candidate who is Muslim. - A candidate who is a Woman. - A candidate who is Mormon.	Carmines & Schmidt, 2020	United States (N = 2338)
Abortion	I have ever had an induced abortion (ended a pregnancy on purpose).	Huber-Krum et al., 2020	Turkey (N = 4040 women)
Belief in human caused climate change	Global warming/climate change is not caused by humans.	Beiser-McGrath & Bernauer, 2021	Germany (N = 3620); United States (N = 3640)

Buying US dollar on the black market (as an illegal currency)	I am going to read you a list of [...] ways in which Argentines regularly relate with the rest of the world. After reading them, I want you to tell me how many of these activities you have undertaken during the last five years. [...] Buy dollars in the parallel market.	Schiumerini & Steinberg, 2020	Argentina (N = 4300)
Sex buying behaviour of adult men	Paid for a sexual encounter from a prostitute.	Roe-Sepowitz et al., 2019	United States (N = 2525)
Perception and experience of electoral manipulation	- I was worried that a family member would lose employment in the public works programme if I voted for the wrong candidate. - I expected a favour from the mayors' men if I voted well. - I worried that I would owe money to my creditor if I voted badly. - During the last elections, I was offered a gift, drink or food in exchange for my vote.	Mares & Young, 2019	Hungarian (N = 1838)
Attitude towards Female genital cutting	How many of these cards show thing that you want for your daughter? [...] Female Circumcision	Gibson et al., 2018	Ethiopia (N = 1620)
Experiences of violence	In the past month (i.e., in the past 30 days), I was beaten, kicked or otherwise physically hurt by my partner. In the past month (i.e., in the past 30 days), my partner forced me to have sex. In the past month (i.e., in the past 30 days), I or another household member have slapped, beaten or kicked my child/my children.	Ebert & Steinert, 2021	Germany (N = 3818 women)
Attitudes towards immigration	How many of the following make you angry? - Immigrants moving into your community. - Candidate or party promises to stop immigration.	Harris et al., 2018	South Africa (N = 1088)

List experiments were also used during the COVID-19 pandemic to collect self-report data from respondents - despite the expected sensitivity of questions on this area of research (Timmons et al., 2021; Larsen et al., 2020; Becher et al., 2020; Munzert & Selb, 2020). A more detailed examination of these case studies and the topic of compliance with COVID-19 regulations itself reveals clear parallels with the topic of corruption.

In both subject areas, behaviours that can be described as sensitive are addressed in surveys. This categorisation can be systematically derived from the dimensions of Tourangeau and Yan (2007) on the basis of various aspects. Firstly, violations of anti-COVID-19 regulations and anti-corruption regulations can be classified as behaviour that can be sanctioned by the state. This means that respondents who have violated these regulations would have to expect negative consequences if the violation becomes public.⁷

Furthermore, violations of both compliance regulations can be regarded as beneficial for a restricted group and detrimental to society at large. In the case of anti-COVID-19 regulations, a violation, e.g., of contact restrictions, poses a risk that could lead to the spread of the virus, higher illnesses rates and a higher burden on the healthcare system. From a sociological perspective, corruption can be understood as deviant behaviour by public officials. This corrupt practice brings advantages for the public official and the clients, and at the same time disadvantages for uninvolved third parties (Graeff & Dombois, 2019, p. 125). In this regard, corruption and violations of COVID-19 regulations are comparable.

This also applies from the perspective of economic theory, which defines corruption as the pursuit of self-interest through actions that result in external costs borne by society at large. Corruption hinders the functioning of the entire economy, slows economic growth and can therefore have a negative impact on society (Beck & Nagel, 2019, p. 41). As these examples show, breaking the compliance rules in both areas are therefore a form of benefit-taking. This behaviour must be kept secret, as both negative legal and negative social sanctions could follow if it is uncovered. In the case of acts of corruption, the threat of disclosure is also increased because the accomplices of the corrupt act expect secrecy.

⁷ It should be briefly mentioned that this applied to surveys that took place during the pandemic and at the same time as such regulations applied. For retrospective surveys, for instance in 2024, it cannot be assumed that there will be pronounced sanctions, as anti-COVID-19 regulations no longer apply. In contrast, regulations against corruption remain relevant.

The previous explanations assume that the sensitivity of the compliance items occurs selectively. Respondents who have adhered to anti-corruption regulations or anti-COVID-19 regulations are unlikely to perceive questions about these as intrusive, threatening or normatively loaded. The sensitivity depending on the correct answer of respondents is therefore also a similarity of both survey topics.

In addition to the aforementioned aspects, there is one final similarity that can be attributed, at least in part, to the preceding points. The compliance regulations in both areas can/could only be partially monitored and enforced. This leads to a substantial *dark field* if looking at official statistics, such as the police crime statistics (for Germany: Police Crime Statistics (PKS) of the Federal Republic, Federal Criminal Police Office). In this sense, both topics require surveys to estimate the prevalence of behaviour.

As already described in chapter 1.1, traditional surveys on sensitive topics can be difficult. For this reason, the Item Count Technique was developed as one specific survey method for sensitive topics. The first article of my dissertation shows that this method has already been used to survey violations of COVID-19-related regulations. In contrast, corruption was a topic that had not yet been investigated with list experiments at the time of publication of the first article of this dissertation. However, it is concluded from the similarities between the survey topics mentioned above that ICT could also be an appropriate method for surveying corruption.

It should be noted that the state of research has progressed since the article was finalised in December 2020. In July 2021, Detkova, Tkachenko and Yakovlev published a study on views on corruption in Russia, which surveyed the opinions of public procurement officials by a list experiment. Their list consists of four or five statements and the respondents are asked to count how many “*create problems for the development of the public procurement system in Russia*” (Detkova et al., 2021, p. 221). Overall, three sensitive aspects were tested. The authors justify this approach with the sensitivity of direct questions on corruption, especially for public officials (Detkova et al., 2021, p. 217; 221).

Another list experiment on the experience of corruption in Romania was conducted at the end of 2018 and the beginning of 2019 (Agerberg, 2022). The objective of the study was to empirically test the hypothesis that items on respondents' personal experience of corruption are associated with a sensitivity bias. The list experiment was employed by Agerberg as a

method of eliminating a sensitivity bias. The results of the list experiment were compared with answers to a directly formulated question. Agerberg also justifies the choice of the list experiment with the sensitivity of the corruption item and then confirms this in the analysis (Agerberg, 2022, p. 670).

Another article by the authors Li and Meng (2020)⁸ uses a list design to investigate the personal experience of corruption among Chinese respondents. Precisely, the sensitive item in their list experiment was whether respondents „*[o]ffered extra expenses to public officials for doing private business*“ (Li & Meng, 2020, p. 440). The authors are therefore asking about the respondents' lifelong experience of corruption. Li and Meng examine the relationship between corruption experience, perceived spread of corruption and assessments of anti-corruption efficacy⁹. They justify the use of the list experiment design with the requirement to measure respondents personal experience with corruption precisely and unbiased (Li & Meng, 2020, p. 432).

The characteristics systematically presented in the first article of my dissertation – the various similarities between corruption items and COVID-19-related compliance items as well as the final endorsement of list experiments for the collection of corruption items – are underpinned by these new publications. It can thus be concluded that the additional financial, planning and organisational effort required for a list experiment is justified if the objective is to survey items pertaining to corrupt behaviour or opinions regarding corruption.¹⁰

⁸ This article was published in 2020. At the time of drafting my book chapter, this paper was not included, as the Journal of Chinese Political Science was not included in the Social Sciences Citation Index (SSCI) until 2021.

⁹ The other variables in the analysis were collected through direct questions. Based on the dimensions of Tourangeau and Yan (2007, see chapter 1.1), the assessments of the prevalence of corruption (in 6 government institutions), the assessment of anti-corruption capacity and the satisfaction with anti-corruption enforcement could also be affected by sensitivity. The authors note this in their presentation of the variables, pointing to non-response rates of 20 to 30 percent (Li & Meng, 2020, p. 442f).

¹⁰ According to Transparency International's Corruption Perceptions Index, Germany might be more of a low prevalence country according to corruption or bribery (Transparency International, 2023). List experiments are also suitable for sensitive items with low prevalence (De Jonge & Nickerson, 2014, p. 659). This can be seen as an additional, more statistical argument for list experiments in research on corruption, even in countries with high compliance.

This first contribution of my dissertation is published in German as a chapter of the book "Corona und Korruption. Gesellschaftswissenschaftliche Analysen" (edited by Sebastian Wolf and Peter Graeff) in May 2022.

I authored this chapter on my own. The original text can be found in chapter 2.2, the English version in chapter 2.1. The basic draft was submitted on 31.12.2020 and the last minor changes were made on 26.11.2021.

1.2.2 Second Paper: Vignette studies

The second article of my dissertation also pursues a theoretical argumentation. As part of a compendium on corruption research, it examines vignette-based factorial surveys as a survey method. Vignette studies are an established method for modelling and collecting individual data.

In a vignette study, respondents are confronted with a short scenario. This scenario describes situations or people. The descriptions contain various aspects, known as vignette dimensions, which characterise the person or situation in a multidimensional way (Atzmüller & Steiner, 2010, p. 128). There are different variants, called levels, for each vignette dimension. Following a vignette, respondents are asked for a personal assessment of the scenario. Depending on the design, respondents are randomly shown one or more vignettes. In the case of several vignettes, respondents answer the same question(s) after every vignette.

Different respondents are confronted with vignettes that contain different characteristics of the vignette dimensions. This procedure can be explained using an example. In their study, Wolter, Raviv and Mertens (2023) showed German respondents variations of the following vignette:

Table 5: Example Vignette

This residential location would *not change your current financial housing costs* (rent, utilities, interest, loan, maintenance costs). There are *many foreigners* living in the residential area. There is also *an active Muslim* community there. Furthermore, it is known that *many poor people* live there. Looking around the neighbourhood, you can see that the streetscape looks *rather run-down and untidy*. An environmental tax for *global climate protection projects* is *financed differently for this residential area and is not payable*.

In general terms, how attractive do you personally find this residence overall?

Very unattractive										Very attractive
1	2	3	4	5	6	7	8	9		10

Source: Wolter et al., 2023, p. 273

The example presented here describes a residential region and requests respondents to rate the attractiveness of the residence in question. The words in *italics* represent the seven vignette dimensions. Each of these dimensions had either three or four characteristics, one of which is shown in the example. Each respondent was presented seven variants of this vignette to assess the attractiveness of the residence in question (Wolter et al., 2023, p. 273f).

The variation of the levels in a factorial survey is carried out experimentally. The resulting data allows the separate analysis of individual dimensions as well as the calculation of interactions between dimensions. The use of vignettes also requires respondents to consider a number of factors simultaneously. This aspect renders vignettes a more realistic question format than a series of direct questions (Dülmer, 2016, p. 3; Sauer et al., 2020, p. 197).

Furthermore, vignettes can be integrated into surveys with large numbers of cases. A factorial survey can be implemented as a "*large-scale population-based experiment*" (Mutz, 2011, p. 54). In this case, vignette-based factorial surveys entail a combination of an experimental design and a survey with an extensive sample. The characteristics of the vignette dimensions vary experimentally. At the same time vignettes can be applied in larger samples than classic (field or laboratory) experiments. This combination leads to advantages resulting from the experimental setting and the (possibly representative) sample. In particular, vignette studies offer high external as well as internal validity (Aguinis & Bradley, 2014, p. 366; Mutz, 2011, p. 5; Schlüter & Schmidt, 2010, p. 93).

In addition to these advantages, which are already described in the literature, vignettes can also be described as an indirect survey method (Sauer et al., 2020, p. 197; Wallander, 2009, p. 506). Respondents provide information about their opinions, attitudes and behavioural intentions. Nevertheless, this is conducted as an evaluation of a hypothetical scenario or assessment of a fictitious person. This approach is often cited in the literature as a way of mitigating social desirability bias (Wallander, 2009, p. 506; Auspurg et al., 2014, p. 137).

Vignette-based factorial surveys are becoming increasingly popular in social science research (Treischl & Wolbring, 2022, p. 142). Vignettes have rarely been explicitly linked to the

phenomenon of sensitivity. At the same time, vignettes have often been used for topics that can be characterised as sensitive. This can be derived indirectly from two comprehensive reviews on the use of factorial surveys in sociological research.

The first study by Wallander (2009) analysed the period from 1982¹¹ to 2006. A total of 106 articles from sociological journals that reported on factorial surveys were identified. Of particular interest in this publication is the examination of the subject areas in which factorial surveys were used. The following text excerpt illustrates overlaps in the use of factorial surveys with subject areas that can be characterised as sensitive based on the dimensions of sensitivity by Tourangeau and Yan (2007, see chapter 1.1):

"[A] large proportion of the publications reviewed focused on subjects related to the field of crime, law and deviance. Within this broad area, the approach has frequently been used in the study of [...] public punishment preferences [...], sexual offending [...], and corporate crime [...]. The subdiscipline category labelled family and social welfare includes a number of studies investigating the judgements of welfare professionals, most of them in relation to child abuse [...]. The focus of the studies classified in the health and medicine category varies, with one study focusing on popular conceptions of mental illness [...] for example, and another comparing the disease-related stigma of AIDS and cancer [...]" (Wallander, 2009, p. 508f).

A more recent meta-analysis by Treischl and Wolbring (2022) covers 441 publications from 1982 to 2018. The findings of this meta-analysis indicate a comparable pattern. The authors initially note a significant increase in publications with factorial surveys (Treischl & Wolbring, 2022, p. 142f). Overall, factorial surveys continue to be used most frequently on *"Crime and Justice"* topics (Treischl & Wolbring, 2022, p. 149f). A consideration of the dimensions of sensitivity as defined by Tourangeau and Yan (2007, see chapter 1.1) reveals that criminal offences must be characterised as sensitive by definition. Criminal offences are linked to a threat of disclosure and respondents may be subject to sanctions if they consent to corresponding behavioural items in a survey. Furthermore, opinions on criminal offences and their sanctioning could also be affected by social desirability and therefore are sensitive.

¹¹ 1982 is the first year included due to an influential publication on the method by Rossi and Nock (1982).

Looking at the list of all publications¹² from the review by Treischl and Wolbring (2022), three studies are of particular interest for this work. The three studies deal with the topic of corruption (Carreras & Vera, 2018; Banerjee et al., 2014; Eggers et al., 2018). The authors examine how the descriptions of politicians whose past behaviour is described as corrupt to varying degrees affect the willingness to participate in elections (Carreras & Vera, 2018) or might affect voting decisions in favour of the described candidates (Banerjee et al., 2014; Eggers et al., 2018). These three studies deal with respondents' opinions and behavioural intentions in response to the corruption of protagonists. The authors justify the choice of the vignette-based factorial survey with the modelling possibilities and the separation of dimensions that are combined in reality (Carreras & Vera, 2018, p. 81; Banerjee et al., 2014, p. 395). Nevertheless, the three studies serve as illustrative examples of the use of this survey paradigm for topics that can be described as sensitive according to the dimensions proposed by Tourangeau and Yan (2007, see chapter 1.1).

In the second article of my dissertation, we systematically present the extent to which vignettes are suitable for collecting data on sensitive topics. This is systematically explained in the article using the specific topic of corruption. As in the first article, the three dimensions of sensitivity according to Tourangeau and Yan (2007) are used. In contrast to list experiments (see chapter 1.2.1), vignettes have already been used to collect data on selected phenomena in the field of corruption research (further examples: Vera, 2020; Dickel & Graeff, 2018; Kleinewiese & Graeff, 2020; Graeff et al., 2014). Our article therefore does not argue in favour of using vignettes for the first time in the field of corruption research, but in favour of expanding the use of this survey paradigm.

The illustrative studies described in the article underline the increasing utilisation of this survey paradigm in corruption research. In addition, the description of the method already highlights the advantages that apply to corruption research if vignettes are used. Vignette-based factorial surveys enable the experimental variation of different factors that could influence attitudes and behavioural intentions with regard to corruption. In the case of corruption - as a sensitive topic - the description of a scenario also has the advantage of creating a certain distance. The respondents evaluate the scenarios presented to them, rather

¹² I would like to express my gratitude to Dr. Edgar Treischl, who provided this comprehensive list of all 441 studies upon request.

than reporting on their own past behaviour. This can result in a decrease in social desirability bias (Wallander, 2009, p. 506). It can also be assumed that the sensitivity bias is reduced, because the other two dimensions of sensitivity according to Tourangeau and Yan (2007, see chapter 1.1) - intrusiveness and threat of disclosure - are weakened or even completely avoided by the indirect questioning technique as well.

This second article paper is currently being under review as a part of the book "Handbuch Korruptionsforschung" (edited by Sebastian Wolf and Olaf Meyer). This chapter was a joint contribution with Professor Dr. Peter Graeff.

The Declaration about the Co-Authorship and the DFG-Guidelines can be found in chapter 6. The first draft of this publication is written in German. The original text from 04.04.2024 can be found in chapter 3.2, the English transcript in chapter 3.1.

1.2.3 Third Paper: Disentangling Differences in Perceived Sensitivity and Misreporting Behaviour: A Factorial Survey

The third article of my dissertation is an empirical paper that applies a vignette-based factorial survey. With reference to the introductory quote, this part of my dissertation explores the question of "*whether sensitivity bias is likely to be a problem*" (Blair et al., 2020, p. 1297; see chapter 1.1).

Sensitivity research assumes that sensitivity can lead to item non-response as well as misreporting. Therefore, sensitivity research uses specific survey methods (see chapter 1.1). Item non-response and any associated biases can be observed. In contrast, biases are not as readily apparent in the case of misreporting. Misreporting was therefore investigated in various ways in order to analyse the effects of sensitivity on this response behaviour. On the one hand, research often relies on group comparisons and on the other hand, some validation studies can be found.

The studies in the first category compare the answers of respondents depending on the survey technique. For instance, the RRT or ICT (see chapter 1.1) is used for one part of a group and direct questions for the other part of the group. The answers are then compared at the

aggregate level. As the prevalence is not known for many subject areas, the *"more is better"* assumption is often followed (Ehler et al., 2021, p. 6; Krumpal, 2013, p. 2033; Tourangeau & Yan, 2007, p. 863). In the context of item non-response, the method that yields a greater number of responses is deemed to be more favourable. When focusing on the prevalence of a sensitive opinion or behaviour, the higher prevalence estimate is considered less biased and the method that suggests the higher estimate is favoured (e.g., Kissling & Jackson, 2022, p. 37; Lensvelt-Mulders et al., 2005; Cobo et al., 2021, p. 13; Ehler et al., 2021, p. 23). The results of such studies vary widely (for an overview e.g., John et al., 2018, p. 103 or Krumpal, 2013, p. 2038ff). However, one common feature of these studies should be emphasised. The topics are characterised as sensitive by the researchers and it is assumed that the assessment of sensitivity does not differ between researchers and respondents.

The second group of studies are validation studies. It is important to note that these studies are still relatively scarce and limited to specific topics due to the confined availability of validation information (Rosenfeld et al., 2016, p. 783). In validation studies, the respondents' answers are compared with external information. Some examples are validation studies among university graduates (surveyed on their final grades; Kreuter et al., 2008), voters (compared with the results of a referendum; Rosenfeld et al., 2016) or patients of an abortion clinic (asked about abortions and compared with information given by the abortion clinic; Rasinski et al., 1999). With few exceptions, the information is available at the meso or macro level (Kuhn & Vivyan, 2021, p. 382).¹³ Furthermore, these studies also refer to topics that researchers have characterised as sensitive.

In an article by Coutts and Jann (2011) a different approach was employed. The authors also tested the RRT and UCT¹⁴ and compared various parameters for these techniques with direct questions. In addition, they also asked the respondents to assess the sensitivity of six items. From the assessments of *"whether committing the behaviour would be all right and whether it would be uncomfortable for most people to admit to having done so"* (Coutts & Jann, 2011, p. 177) the authors created an index. This *"Perceived Question Sensitivity"* was shown in a

¹³ Some exceptions are validation studies among lung patients about smoking behaviour compared to results of a breath test (Fox & van der Palen, 2013) or only convicted offenders surveyed on delinquent behaviour (Preisendörfer & Wolter, 2014).

¹⁴ As a reminder: UCT and ICT are two alternative terms for list experiments which is the term mostly used in this dissertation.

descriptive overview, but was not systematically included in the analysis of the performance of RRT or UCT (Coutts & Jann, 2011, p. 184).

The third article of my dissertation builds upon the idea of sensitivity assessments conducted by the respondents themselves and also employs a vignette-based factorial survey (see chapter 1.2.2). My vignette design describes different survey situations with varying survey items, bystanders and different survey modes (see chapter 4, appendix A1 for the wording of each vignette level). The respondents are then asked for an assessment of the sensitivity as well as an assessment of the non-response probability and misreporting probability (see Table 6):

Table 6: Example of a vignette presented to respondents, including questions to assess

Please imagine the following situation: You are taking part in an <i>online survey</i>. <i>No other people are present at the time.</i> You are asked the following question: <i>In the last 5 years, have you offered someone money to do something for you that no one else should know about?</i> ☐ Yes ☐ No ☐ Prefer not to say																																											
Would you answer this question? ☐ Yes ☐ No																																											
How uncomfortable is the overall situation described above for you personally? <table> <tr> <td></td> <td>Not uncomfortable at all</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>Very uncomfortable</td> </tr> <tr> <td></td> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> <td>7</td> <td>8</td> <td>9</td> </tr> <tr> <td>I think this situation is ...</td> <td>☐</td> <td>☐</td> <td>☐</td> <td>☐</td> <td>☐</td> <td>☐</td> <td>☐</td> <td>☐</td> <td>☐</td> <td>☐</td> </tr> </table>												Not uncomfortable at all									Very uncomfortable		0	1	2	3	4	5	6	7	8	9	I think this situation is ...	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Not uncomfortable at all									Very uncomfortable																																	
	0	1	2	3	4	5	6	7	8	9																																	
I think this situation is ...	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐																																	
How likely would you answer the question truthfully? <table> <tr> <td></td> <td>Very unlikely</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>Very likely</td> </tr> <tr> <td></td> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> <td>7</td> <td>8</td> <td>9</td> </tr> <tr> <td>It is...</td> <td>☐</td> <td>☐</td> <td>☐</td> <td>☐</td> <td>☐</td> <td>☐</td> <td>☐</td> <td>☐</td> <td>☐</td> <td>☐</td> </tr> </table>												Very unlikely									Very likely		0	1	2	3	4	5	6	7	8	9	It is...	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Very unlikely									Very likely																																	
	0	1	2	3	4	5	6	7	8	9																																	
It is...	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐																																	

Note: This English version is translated from the original German questionnaire. The varying dimensions are shown in *italics*. The survey item in the vignette (in this example: the bribery question) was shown with

In this way, the resulting data set includes assessments of the sensitivity as well as assessments of the potential response behaviour of respondents. Sensitivity is therefore neither defined a priori nor derived from the response behaviour (e.g., the non-response rate).

This hierarchically structured data set at the individual level makes it possible to examine the relationships between sensitivity assessment and estimated response behaviour. The respondents do not give any direct answers to the (potentially) sensitive questions or sensitive answer options. The objective of this design is to minimise biases due to sensitivity (see section 1.2.2).

The article pursues a hypothesis-testing. To this end, a total of six hypotheses on two dependent variables (sensitivity assessment and misreporting probability) are formulated. The postulated relationships were considered separately in the literature and are now estimated simultaneously as part of the vignette-based factorial survey design.

The first dependent variable is the sensitivity assessment. Respondents indicated on a scale of 0 to 9 how uncomfortable they felt about the situation described in the vignette scenario. The influence of the three vignette dimensions “survey item”, “presence of bystander” and “survey mode” were analysed. The hypotheses regarding the influence of bystanders and interviewers/ the survey mode on the sensitivity assessment were not confirmed. Only the different survey items caused significant differences in the respondents' sensitivity assessment.

It is interesting to note that the selected survey items (age, religion, sexual orientation, infidelity, experience of violence and bribery) only partially corresponded to expectations with regard to the relative sensitivity assessments. This emphasises that the sensitivity assessment of researchers (e.g., based on the theoretical dimensions of Tourangeau & Yan, 2007, see chapter 1.1) can deviate from the perception of respondents. This was the rationale for the sensitivity assessments by the respondents (instead of an assessment by me). The approach proved to be accurate in the subsequent analysis.

The second dependent variable of this design was the misreporting probability. Respondents were asked to assess whether they would answer the question truthfully if they were in the

interview situation described in the vignette. The responses were again placed on a scale of 0 to 9. This variable was recoded for the analysis in such a way that a high value indicates a high misreporting probability. Sensitivity assessment and anonymity assessment were identified in the theoretical chapter as core variables for the variation in misreporting probability. In addition to the main effects of these independent variables, an interaction effect was also included. The sensitivity assessment showed a significant and reinforcing influence on the probability of misreporting across all models. The same was initially shown for the anonymity assessment of respondents, which was coded as a violation of anonymity (in preparation for the subsequent interaction term). The more likely respondents expect a violation of anonymity, the higher the probability of misreporting was evaluated. The integration of an interaction effect between these two independent variables proved to be interesting. The interaction term became significant in some cases and the main effect of anonymity violation lost its statistical significance. The main effect of the sensitivity assessment remained significant in the model with the interaction effect. This means that the sensitivity assessment had a reinforcing effect on the misreporting probability *even* under the assumption of anonymous responses.

These results suggest that even anonymity of respondents' answers is not sufficient to prevent deliberate misreporting regarding sensitive topics. In light of the preceding article, further research is required that addresses factors beyond those related to anonymity.

This last article of the dissertation is currently being under review (after revisions) as part of the journal *Quality & Quantity. International Journal of Methodology*. I authored this paper on my own. The revised version of the article (as of 24.04.2024) can be found in chapter 4.

1.3 Relationship of the Papers

Each of the three papers deal with different aspects of the topic of sensitivity. All of them use the three dimensions of sensitivity according to Tourangeau and Yan (2007) as a theoretical perspective.

Chapter 1.1 summarised that the phenomenon of sensitivity is not new and that biases due to sensitivity are assumed. Negative biases in response behaviour were identified for a number of topics and special survey techniques were developed. Based on this, the articles build on the current state of research and focus on selected aspects.

The first and second paper use a theoretical perspective and shed light on how to address the challenges of collecting self-disclosures on the specific topic of (anti-) corruption.

The third article builds on the approach of the second article by conducting a vignette study after the benefits of the approach have been explained in the second article. Additionally, the final article of this dissertation also deals with the effects of sensitivity on response behaviour *if* a direct question format would be chosen (which is explicitly not recommended in articles one and two). The results clearly show that a stronger sensitivity assessment leads to a higher probability of item nonresponse and misreporting. In this way, article three empirically reinforces the theory-based argumentation of the first two articles.

1.4 How the dissertation fills previous research gaps

This dissertation contributes to filling the research gaps in the field of sensitivity research. In particular, it sheds light on two facets.

On the one hand, structured, theory-based reasonings are provided for assessing the sensitivity of direct questions on the topic of corruption (for more details see sections 1.2.1 and 1.2.2). In the course of this articles, two potential strategies for addressing sensitivity (and sensitivity bias) qua design are suggested. These publications thus make it clear how the data collection of opinions on and experiences with corruption can be improved.

This structured approach as well as the two proposed methods can be transferred to other survey topics that may be affected by sensitivity and have not yet been analysed using the two methods. In this way, the two articles provide a framework for further investigation of sensitivity (effects) on a larger scale.

On the other hand, the third paper takes an empirical approach and examines factors influencing the sensitivity assessment as well as the connection between sensitivity assessment and anonymity assessment with the response behaviour of respondents. It is

shown that the sensitivity assessment was strongly influenced by the directly formulated items and that this sensitivity assessment also had a significant influence on the conscious response behaviour of respondents. This influence could not be completely eliminated even by the anonymity of the responses. This outcome is particularly noteworthy in light of the prevailing notion that improved anonymity will enhance response rates and data quality.

This aspect of the dissertation also provides a number of valuable points of reference for further research. The results showed different sensitivity assessments from respondents which did not always match my theory-driven expectations. This leads to the conclusion that the sensitivity assessment of researchers and respondents *can* differ. An assessment by researchers regarding the sensitivity of topics can therefore be incomplete and in the future, it seems promising to ask respondents about their assessment of sensitivity.

This leads to another avenue for further research, which also illustrates the contribution of my third article. In the third paper of my dissertation, I developed a factorial survey design that measures the sensitivity of a total of six items in a novel way. The six items on demographics and behaviour have not tested for sensitivity before. This method can be extended to other items in the future, for instance to questions on socially (un)desirable opinions and attitudes.

1.5 References

- Agerberg, M. (2022). Corrupted Estimates? Response Bias in Citizen Surveys on Corruption. *Political Behavior*, 44, 653-678. doi:10.1007/s11109-020-09630-5.
- Aguinis, H., & Bradley, K. J. (2014). Best Practics Recommendations for Designing and Implementing Experimental Vignette Methodolody Studies. *Organizational Research Methods*, 17(4), 351-371. doi:10.1177/1094428114547952.
- Atzmüller, C., & Steiner, P. M. (2010). Experimental Vignette Studies in Survey Research. *Methodology*, 6(3), 128–138. doi:10.1027/1614-2241/a000014.
- Auspurg, K., Hinz, T., Liebig, S., & Sauer, C. (2014). The Factorial Survey as a Method for Measuring Sensitive Issues. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, & P. Sturgis,

- Improving Survey Methods (p. 137-149). New York: Routledge.
doi:10.4324/9781315756288.
- Banerjee, A., Green, D. P., McManus, J., & Pande, R. (2014). Are Poor Voters Indifferent to Whether Elected Leaders Are Criminal or Corrupt? A Vignette Experiment in Rural India. *Political Communication*, 31(3), 391-407. doi:10.1080/10584609.2014.914615.
- Becher, M., Stegmueller, D., Brouardc, S., & Kerrouhec, E. (2020). Comparative experimental evidence on compliance with social distancing during the COVID-19 pandemic. medRxiv. the Preprint Server for Health Sciences.
doi:10.1101/2020.07.29.20164806.
- Beck, L., & Nagel, V. (2019). Korruption aus ökonomischer Perspektive. In P. Graeff & T. Rabl (Hrsg.), *Was ist Korruption? Begriffe, Grundlagen und Perspektiven gesellschaftswissenschaftlicher Korruptionsforschung* (S. 39–58). Nomos.
- Beiser-McGrath, L. F., & Bernauer, T. (2021). Current surveys may underestimate climate change skepticism evidence from list experiments in Germany and the USA. *PLoS ONE*, 16(7), e0251034. doi:10.1371/journal.pone.0251034.
- Blair, G., Coppock, A., & Moor, M. (2020). When to Worry about Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments. *American Political Science Review*, 114(4), 1297–1315. doi:10.1017/S0003055420000374.
- Carmines, E. G., & Schmidt, E. R. (2020). Prejudice and Tolerance in US Presidential Politics: Evidence from Eight List Experiments in 2008 and 2012. *Political Science & Politics*, 54, 24-32. doi:10.1017/S1049096520000876.
- Carreras, M., & Vera, S. (2018). Do Corrupt Politicians Mobilize or Demobilize Voters? A Vignette Experiment in Colombia. *Latin American Politics and Society*, 60(3), 77–95. doi:10.1017/lap.2018.25.
- Celhay, P., Meyer, B. D., & Mittag, N. (2022). What Leads to Measurement Errors? Evidence from Reports of Program Participation in Three Surveys. *IZA Discussion Papers*.
- Chapkovski, P., & Schaub, M. (2022). Solid support or secret dissent? A list experiment on preference falsification during the Russian war against Ukraine. *Research and Politics*, 1-6. doi:10.1177/20531680221108328.

- Cobo, B., Casillo, E., López-Torrecillas, F., & Del Mar Rueda, M. (2021). Indirect questioning methods for sensitive survey questions: Modelling criminal behaviours among a prison population. *PLoS ONE*, 16(1), e0245550. doi:10.1371/journal.pone.0245550.
- Coutts, E., & Jann, B. (2011). Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT). *Sociological Methods & Research*, 40, 169-193. doi:10.1177/0049124110390768.
- De Jonge, C. P., & Nickerson, D. W. (2014). Artificial Inflation or Deflation? Assessing the Item Count Technique in Comparative Surveys. *Political Behavior*, 36, 659-682. doi:10.1007/s11109-013-9249-x.
- Detkova, P., Tkachenko, A., & Yakovlev, A. (2021). Gender heterogeneity of bureaucrats in attitude to corruption: Evidence from list experiment. *Journal of Economic Behavior and Organization*, 189, 217-233. doi:10.1016/j.jebo.2021.06.034.
- Dickel, P., & Graeff, P. (2018). Entrepreneurs' propensity for corruption: A vignette-based factorial survey. *Journal of Business Research*, 89, 77-86. doi:10.3109/10826080009148424.
- Droitcour, J. A., Caspar, R., Hubbard, M., Parsely, T., Visscher, W., & Ezzati, T. (1991). The item count technique as a method of indirect questioning: a review of its development and a case study application. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, & S. Sudman, *Measurement Errors in Surveys* (p. 185-210). New York: Wiley.
- Dülmer, H. (2016). The Factorial Survey: Design Selection and its Impact on Reliability and Internal Validity. *Sociological Methods & Research*, 1-44. doi:10.1177/0049124115582269.
- Ebert, C., & Steinert, J. (2021). Prevalence and risk factors of violence against women and children during COVID-19, Germany. *Bull World Health Organ*, 99, 429–438. doi:10.2471/BLT.20.270983.

- Eggers, A., Vivyan, N., & Wagner, M. (2018). Corruption, Accountability, and Gender: Do Female Politicians Face Higher Standards in Public Life? *The Journal of Politics*, 80(1), 321–326. doi:10.1086/694649.
- Ehler, I., Wolter, F., & Junkermann, J. (2021). Sensitive Questions in Surveys. A comprehensive meta-analysis of experimental survey studies on the performance of the item count technique. *Public Opinion Quarterly*, 85(1), 6-27. doi:10.1093/poq/nfab002.
- European Social Survey. (2020). ESS Round 10 Source Questionnaire. London. <https://ess.sikt.no/en/study/172ac431-2a06-41df-9dab-c1fd8f3877e7/438>.
- Fox, J. A. (2016). *Randomized response and related methods: surveying sensitive data*. Los Angeles, London, New Delhi, Singapore, Washington, DC: Sage.
- Fox, J. A., & van der Palen, J. (2013). Mixture randomized item-response modeling: a smoking behavior validation study. *Statistics in Medicine*, 32(27), 4821-4837. doi:10.1002/sim.5859.
- GESIS Leibniz-Institut für Sozialwissenschaften. (2023). German General Social Survey - ALLBUS 2021. Questionnaire ZA5280. Cologne. https://search.gesis.org/research_data/ZA5282.
- Gibson, M. A., Gurmu, E., Cobo, B., Rueda, M. M., & Scott, I. M. (2018). Indirect questioning method reveals hidden support for female genital cutting in South Central Ethiopia. *PLoS ONE*, 13(5), e0193985. doi:10.1371/journal.pone.0193985.
- Glynn, A. N. (2013). What can we learn with statistical truth serum? Design and Analysis of the List Experiment. *The Public Opinion Quarterly*, 77, 159-172. doi:10.1093/poq/nfs070.
- Graeff, P., & Dombois, R. (2019). Soziologische Zugänge zur Korruptionsproblematik. In P. Graeff, & T. Rabl, *Was ist Korruption? Begriffe, Grundlagen und Perspektiven gesellschaftswissenschaftlicher Korruptionsforschung* (2nd edition., p. 123-142). Baden-Baden: Nomos. doi:10.5771/9783845289847.

- Graeff, P., Sattler, S., Mehlkop, G., & Sauer, C. (2014). Incentives and Inhibitors of Abusing Academic Positions: Analysing University Students' Decisions about Bribing Academic Staff. *European Sociological Review*, 30(2), 230-241. doi:10.1093/esr/jct036.
- Greenberg, B. G., Abul-El, A.-L. A., Simmons, W. R., & Horvitz, D. G. (1969). The Unrelated Question Randomized Response Model: Theoretical Framework. *Journal of the American Statistical Association*, 64(326), 520-539.
- Harris, A. S., Findley, M. G., Nielson, D. L., & Noyes, K. L. (2018). The Economic Roots of Anti-immigrant Prejudice in the Global South: Evidence from South Africa. *Political Research Quarterly*, 71(1), 228-241. doi:10.1177/10659129177340.
- Heimann, D. (2022). Statistical Consideration. *Bright and Dark Fields of Crime*. In R. Heimann, & J. Fritzsche, *Violence Prevention in Education, School, and Club* (p. 21-32). Wiesbaden: Springer. doi:10.1007/978-3-658-38551-4_2.
- Hinsley, A., Hu, S., Chen, H., Garshelis, D., Hoffmann, M., Lee, T. M., . . . Milner-Gulland, E. J. (2021). Combining data from consumers and traditional medicine practitioners to provide a more complete picture of Chinese bear bile markets. *People and Nature*, 3, 1064-1077. doi:10.1002/pan3.10249.
- Huber-Krum, S., Karadon, D., Kurutas, S., Rohr, J., Baykal, S. S., Okcuoglu, B. A., . . . Shah, I. (2020). Estimating abortion prevalence and understanding perspectives of community leaders and providers: Results from a mixed-method study in Istanbul, Turkey. *Women's Health*, 16, 1-13. doi:10.1177/1745506520953353.
- Imai, K. (2011). Multivariate Regression Analysis for the Item Count Technique. *Journal of the American Statistical Association*, 106(494), 407-416. doi:10.1198/jaja.2011.ap10415.
- Iraman, E., Ono, Y., & Kakinaka, M. (2022). Tax compliance and social desirability bias of taxpayers: experimental evidence from Indonesia. *Journal of Public Policy*, 42, 92–109. doi:10.1017/S0143814X21000040.
- ISSP Research Group. (2020). *International Social Survey Programme: Religion IV - ISSP 2018*. Questionnaire ZA7570. Cologne. doi:10.4232/1.13629.

- John, L. K., Loewenstein, G., Acquisti, A., & Vosgerau, J. (2018). When and why randomized response techniques (fail to) elicit the truth. *Organizational Behavior and Human Decision Processes*, 148, 101-123. doi:10.1016/j.obhdp.2018.07.004.
- Kissling, A., & Jackson, H. M. (2022). Estimating Prevalence of Abortion Using List Experiments: Findings from a Survey of Women in Delaware and Maryland. *Women's Health Issues*, 32(1), 33-40. doi:10.1016/j.whi.2021.08.003.
- Kleinewiese, J., & Graeff, P. (2020). Ethical Decisions between the Conflicting Priorities of Legality and Group Loyalty: Scrutinizing the “Code of Silence” among Volunteer Firefighters with a Vignette-Based Factorial Survey. *Deviant Behaviour*, 42(10), 1228-1241. doi:10.1080/01639625.2020.1738640.
- Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social Desirability Bias in CATI, IVR, and Web Surveys. The effects of mode and question sensitivity. *Public Opinion Quarterly*, 72(5), 847-864. doi:10.1093/poq/nfn063.
- Krosnick, J. A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, 5, 213-236.
- Krosnick, J. A. (1999). Survey Research. *Annual Review of Psychology*, 50, 537-567.
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, 47, 2025–2047. doi:10.1007/s11135-011-9640-9.
- Kuhn, P. M., & Vivyan, N. (2021). The Misreporting Trade-Off Between List Experiments and Direct Questions in Practice: Partition Validation Evidence from Two Countries. *Political Analysis*, 30(3), 381- 402. doi:10.1017/pan.2021.10.
- Larsen, M. V., Nyrup, J., & Petersen, M. B. (2020). Do survey estimates of the public’s compliance with COVID-19 Regulations suffer from social desirability bias? *Journal of Behavioral Public Administration*, 3(2), 1-9. doi:10.30636/jbpa.32.164.
- Lensvelt-Mulders, G. J., Hox, J. J., van der Heijden, P. G., & Maas, C. J. (2005). Meta-Analysis of Randomized Response Research. Thirty-Five Years of Validation. *Sociological Methods & Research*, 33, 319-348.

- Li, H., & Meng, T. (2020). Corruption Experience and Public Perceptions of Anti-corruption Crackdowns: Experimental Evidence from China. *Journal of Chinese Political Science*, 25, 431-456. doi:10.1007/s11366-020-09672-w.
- Mares, I., & Young, L. (2019). Varieties of Clientelism in Hungarian Elections. *Comparative Politics*, 51(3), 449-471.
- Miller, J. D. (1984). A new survey technique for studying deviant behavior. The George Washington University. PhD diss.
- Munzert, S., & Selb, P. (2020). Can we directly survey adherence to non-pharmaceutical interventions? Evidence from a list experiment conducted in Germany during the early corona pandemic. *Survey Research Methods*, 14(2), 205-209. doi:10.18148/srm/2020.v14i2.7759.
- Mutz, D. C. (2011). *Population-Based Survey Experiments*. Princeton: Princeton University Press. doi:10.1515/9781400840489.
- National Opinion Research Center. (2023). *The General Social Survey*. Chicago. <https://gss.norc.org/get-documentation/questionnaires>.
- Preisendörfer, P., & Wolter, F. (2014). Who Is Telling the Truth? A Validation Study on Determinants of Response Behavior in Surveys. *Public Opinion Quarterly*, 78(1), 126-146. doi:10.1093/poq/nft079.
- Rasinski, K. A., Willis, G. B., Baldwin, A. K., Yeh, W., & Lee, L. (1999). Methods of Data Collection, Perceptions of Risks and Losses, and Motivation to Give Truthful Answers to Sensitive Survey Questions. *Applied cognitive Psychology*, 13, 465-484.
- Roe-Sepowitz, D., Bontrager, S., Pickett, J. T., & Kosloski, A. E. (2019). Estimating the sex buying behavior of adult males in the United States: List experiment and direct question estimates. *Journal of Criminal Justice*, 63, 41-48. doi:10.1016/j.jcrimjus.2019.04.005.
- Rosenfeld, B., Imai, K., & Shapiro, J. N. (2016). An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions. *American Journal of Political Science*, 60(3), 783-802. doi:10.1111/ajps.12205.

- Rossi, P. H., & Nock, S. S. (1982). *Measuring Social Judgments: The Factorial Survey Approach*. Sage Publications.
- Sauer, C., Auspurg, K., & Hinz, T. (2020). Designing Multi-Factorial Survey Experiments: Effects of Presentation Style (Text or Table), Answering Scales, and Vignette Order. *methods, data, analyses*, 14(2), 195-214. doi:10.12758/mda.2020.06.
- Schiumerini, L., & Steinberg, D. A. (2020). The Black Market Blues: The Political Costs of Illicit Currency Markets. *The Journal of Politics*, 82(4), 1217-1230. doi:10.1086/708242.
- Schlüter, E., & Schmidt, P. (2010). Special issue: Survey experiments. *Methodology European Journal of Research Methods for the Behavioral and Social Sciences*, 6, 93-54. doi:10.1027/1614-2241/a000010.
- Sniderman, P. M., Tetlock, P. E., & Piazza, T. (1992). *Codebook for the 1991 National Race and Politics Survey*. Berkely: Survey Research Center. <https://sda.berkeley.edu/D3/Natlrace/Doc/nrac.htm>.
- Timmons, S., McGinnity, F., Belton, C., Barjaková, M., & Lunn, P. (2021). It depends on how you ask: measuring bias in population surveys of compliance with COVID-19 public health guidance. *Epidemiol Community Health*, 75, 387-387. doi:10.1136/jech-2020-215256.
- Tourangeau, R. (2018). Confidentiality, Privacy, and Anonymity. In D. L. Vannette, & J. A. Krohn, *The Palgrave Handbook of Survey Research* (p. 501-507). palgrave macmillan. doi:10.1007/978-3-319-54395-6_58.
- Tourangeau, R., & Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin*, 133(5), 859–883. doi:10.1037/0033-2909.133.5.859.
- Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Transparency International. (2023). *Corruption Perceptions Index*. <https://www.transparency.org/en/cpi/2023/index/deu>.

- Trappmann, M., Krumpal, I., Kirchner, A., & Jann, B. (2014). Item Sum: A new Technique for Asking quantitative sensitive Questions. *Journal of Survey Statistics and Methodology*, 2, 58-77. doi:10.1093/jssam/smt019.
- Traunmüller, R., Kijewski, S., & Freitag, M. (2019). The Silent Victims of Sexual Violence during War: Evidence from a List Experiment in Sri Lanka. *Journal of Conflict Resolution*, 63(9), 2015-2042. doi:10.1177/0022002719828053.
- Treischl, E., & Wolbring, T. (2022). The Past, Present and Future of Factorial Survey Experiments: A Review for the Social Sciences. *methods, data, analyses*, 16(2), 141-170. doi:10.12758/mda.2021.07.
- Vera, S. B. (2020). Accepting or Resisting? Citizen Responses to Corruption Across Varying Levels of Competence and Corruption Prevalence. *Political Studies*, 68(3), 653-670. doi:10.1177/0032321719868210.
- Wallander, L. (2009). 25 years of factorial surveys in sociology: A review. *Social Science Research* (38), 505-520. doi:10.1016/j.ssresearch.2009.03.004.
- Warner, S. L. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, 60(309), 63-69.
- Wolter, F., & Preisendörfer, P. (2013). Asking Sensitive Questions: An Evaluation of the Randomized Response Technique Versus Direct Questioning Using Individual Validation Data. *Sociological Methods & Research*, 42(3), 321-353. doi:10.1177/0049124113500474.
- Wolter, F., Raviv, O. C., & Mertens, M. (2023). Discriminatory Residential Preferences in Germany - A Vignette Study. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 75, 263-288. doi:10.1007/s11577-023-00906-2.
- World Values Survey Association. (2022). 2017 -2021 World Values Survey Wave 7. Master Survey Questionnaire. <https://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp>.
- Xu, C., & Liu, Y. (2021). Does mislabeling COVID-19 elicit the perception of threat and reduce blame? *Journal of Behavioral Public Administration*, 4(2), 1-13. doi:10.30636/jbpa.42.225.

Yan, T. (2021). Consequences of Asking Sensitive Questions in Surveys. *Annual Review of Statistics and Its Application*, 8, 109–27. doi:10.1146/annurev-statistics-040720-033353.

2 FIRST PAPER

2.1 English Transcript

Corruption and corona. Possible similarities between two survey topics

Introduction

"Corruption is by no means a mainstream topic in national surveys..."

(Richards, 2017)

Researching and uncovering corruption has always been challenging for researchers. The existing pitfalls to researching and uncovering corruption, such as a lack of knowledge about the extent of corruption in a society or company, have been exacerbated to some extent by the COVID-19 pandemic. At the same time, from the perspective of survey research, important lessons for the data collection on the topic of corruption can also be derived from surveys conducted during the pandemic. Some surveys related to the COVID-19 pandemic use questioning techniques that could also be promising and worthwhile for collecting data in surveys on corrupt behaviour. In this article, this potential will be derived from the comparison of behaviours that violate compliance in both areas.

Similar to the understanding of the term in other subject areas, compliance in relation to corruption should be understood as adherence to rules, regulations and standards (Graeff & Steßl, 2017). These regulations are directed against corruption in its various facets. In this context, the term corruption can be understood as the behaviour of actors, more precisely a cooperative and deviant act between at least two individuals. The cooperation becomes deviant behaviour when participants act as representatives of a certain office and at the same time deviate from the role obligations of this office (Graeff & Dombois, 2019). Furthermore, the cooperation can be disadvantageous for uninvolved third parties (Beck & Nagel, 2019). Prominent examples of acts of corruption include the bribery of political office holders, the

circumvention of public tenders or the exchange of mutual favours in a professional context. At its core, corruption-related compliance should therefore be understood as the behaviour of actors who (actively) decide against participating in corrupt behaviour in accordance with a social norm or a law.

Collecting data on the extent of compliance or corrupt behaviour in a company or country is difficult for various reasons. Directly asking about corrupt behaviour in a survey is problematic, because corruption implies secrecy towards uninvolved third parties and is a socially undesirable act. As a result, surveys rarely ask respondents specific questions about corrupt behaviour. Items to assess the extent of corruption that do not directly inquire about the respondent's behaviour are more common (Richards, 2017). Respondents are asked, for instance, to assess the extent of corruption within their country. This article presents an opportunity to show how – despite existing pitfalls – it is also possible to ask respondents specific questions about corrupt behaviour.

This opportunity is derived from procedures for assessing compliance in relation to COVID-19. With regard to the COVID-19 pandemic, the keyword compliance refers to adherence to preventive, non-medical regulations to decrease the spread of the infection (e.g., regular hand washing, wearing face masks, social distancing, contact restrictions). This compliance is considered an important component of pandemic control (Anderson et al., 2020; Betsch, 2020). Nevertheless, it is challenging to assess compliance with these regulations in an objective manner, particularly given that they also pertain to citizens' private spaces (Timmons et al., 2020a). Larsen et al. (2020), on the one hand, state that mobile phone data can be used to map the movement patterns of the population and thus estimate compliance with lockdowns or other regulations. On the other hand, this data does not provide any information about the specific behaviour of individuals when they meet each other. It cannot be concluded from the movement data whether individuals hug each other, shake hands in greeting or consistently wear masks when socialising outside their own household. Furthermore, the behaviour of individuals in their own households, such as washing their hands when returning to their own household, cannot be determined from this data. In order to obtain information on this type of compliance, data from population surveys is required.

For this reason, such questions have also become part of various survey instruments that are intended to provide a data base regarding the extent of various behaviours. In this context,

specific survey techniques from the field of sensitivity research were used which are presented in the course of this article (Becher et al., 2020; Larsen et al., 2020; Timmons et al., 2020b). Subsequently, similarities between corruption-related items and COVID-19-related survey questions will be stated to derive potentials for surveying corrupt behaviour, intentions and tendencies.

Survey research on the COVID-19 pandemic

Variety of topics

Since the onset of the COVID-19 pandemic in spring 2020 numerous surveys have been conducted with reference to this exceptional situation. The items in these surveys cover a broad range and will be summarised here. A total of 6 categories can be identified:

- 1 Items on various general survey topics that were already asked before the COVID-19 pandemic;
- 2 Items about infections with COVID-19;
- 3 Items about the (potential) effects of the COVID-19 pandemic;
- 4 Items on *opinions* regarding (political) regulations to combat the pandemic (excluding medical regulations);
- 5 Items on *compliance* with (political) regulations to combat the pandemic (excluding medical regulations);
- 6 Items about (potential) vaccines and vaccination readiness.

Questions in the first category are congruent with items that were already asked before the pandemic. These items are considered in terms of potential changes that are attributed to the pandemic, specific political decisions, contact restrictions or (various stages of) lockdowns. For instance, Sibley et al. (2020) focus on the effects of the pandemic on respondents' life satisfaction, health and institutional trust. Perrotta et al. (2020) still consider trust in various institutions and the assessment of reliability of various sources of information (e.g., newspapers, social media, radio, stories from friends or acquaintances). In addition, regular studies such as the Socio-Economic Panel (SOEP) are continued in order to record changes in various variables (Kühne et al., 2020).

Questions in the second category relate to (current and past) infections with COVID-19, potential disease progression if respondents were infected and the incidence of infection in

the personal environment of respondents (Kapteyn et al., 2020; Kreuter et al., 2020). Kreuter et al. (2020), for instance, focus on these topics in their surveys and argue that their data collection is faster compared to official statistics. The authors developed a survey tool that reached respondents in more than 200 territories every day via Facebook. Furthermore, the collaboration with Facebook enabled a specific weighting of the data. In this way, the authors collected data that in some cases facilitate more accurate information about the current incidence of infection than official statistics can provide. Furthermore, Post et al. (2020) deal with possible sources of error in the data collection of official statistics and subsequently confirm the collection of survey data to improve the monitoring of the infection incidence.

Questions in the third category focus on the various effects of the COVID-19 pandemic and the regulations to combat the pandemic. On the one hand, items in this category ask about the respondent's assessment of the expected effects in the future, for instance with regard to economic or personal consequences (Juhl et al., 2020). On the other hand, items in this category also refer to changes in everyday life at the time of the survey. These items aiming at potential changes in family life (due to closures of daycare centres or schools) or working life (due to contact restrictions, working from home, short-time working, etc.) (Hipp et al., 2020; Huebner et al., 2020; Kapteyn et al., 2020; Kühne et al., 2020). Other questions of this category focus on the emotional, mental and psychological state of respondents (Hipp et al., 2020; Kreuter et al., 2020; Kühne et al., 2020; Naumann et al., 2020; Perrotta et al., 2020).

Further questions – which are summarised as the fourth category – relate to the understanding, acceptance and approval of the (political) regulations combating the pandemic at the time of the survey. The objective of items in this category is to gain insight into the general evaluation of decisions and their subsequent communication to citizens. Responses to this question category are therefore also of considerable importance to political decision-makers and are collected, for instance, in weekly surveys as part of the Mannheim Corona Study (Juhl et al., 2020).

Questions in the fifth category, which is central to this article, focus on *compliance with* various (political) regulations to combat the pandemic. These questions go beyond the respondents' *opinions* and refer to past or future behaviour in more detail. For instance, questions about contact with friends or relatives despite a requirement to restrict contact (Daoust et al., 2020; Kreuter et al., 2020; Munzert & Selb, 2020; Perrotta et al., 2020) or items

regarding the hand hygiene of respondents (Kapteyn et al., 2020; Utych & Fowler, 2020) are assigned to this category. Furthermore, questions about wearing masks, which are also categorised as a preventive regulation, can also be assigned to this group of questions (Kapteyn et al., 2020). The surveyed behaviours, which are referred to as "*non-pharmaceutical interventions*", are considered preventive and were the only way to contain the pandemic at the beginning (Anderson et al., 2020; Betsch, 2020).

In contrast, combating the pandemic with vaccination is the second important component of crisis management. The items in the sixth category relate to this very possibility. Various items on vaccines as well as items on the willingness of respondents to get vaccinated are assigned to this category (Graeber et al., 2020a, b; Kapteyn et al., 2020; Neumann-Böhme et al., 2020).

Sensitivity

The previous chapter illustrates the range of survey items and topics used in relation to COVID-19. From the perspective of survey research, selected categories are particularly interesting because in some research areas, the pathways of traditional survey techniques were left behind. In the following, I will focus in particular on the survey items regarding *compliance* with regulations to minimise the risk of infection with COVID-19. The surveyed behaviours, such as hand hygiene or contact restrictions, are of central importance for the progress of the pandemic, both in the absence of vaccines and during vaccination campaigns. They are characterised by special features from the perspective of survey research.

One of these characteristics is the sensitivity of the topics relating to compliance. The sensitivity of items leads to (potentially) biased data if respondents do not answer these items or do not answer them truthfully. Following Tourangeau and Yan (2007), three dimensions of sensitivity can be distinguished.

Initially, respondents may perceive items as intrusive and inappropriate for the survey situation. This is especially the case when questions touch upon social taboos. Respondents may then see the content of the question as a social taboo that they generally do not want to talk about. As a result, respondents sometimes refuse to answer which can bias the data (Jann et al., 2019; Tourangeau & Yan, 2007).

Furthermore, the sensitivity of items can be associated with a threat of disclosure or discovery of responses. Some responses to specific questions that could be used as evidence in a criminal case (and for which one could be prosecuted long after the offence has been committed) may therefore be perceived as sensitive by respondents. The degree of sensitivity for a question varies between individual respondents and depends on their potential answer (Tourangeau & Yan, 2007). Respondents whose true answer is sensitive in this sense may refuse to answer. If an item has only one sensitive response, a refusal to answer can be seen as a kind of admission of guilt or suggestion of this sensitive response (Tourangeau & Yan, 2007; Wolter, 2019). This is why respondents might choose an answer option that does not correspond to the truth (Böckenholt, 2014). For instance, respondents who have committed a criminal offence could conceal this in a survey (Skarbek-Kozietulska et al., 2012). This dimension of sensitivity can also lead to biased data, whereby the bias goes beyond item non-response and the responses can also result in a biased description of a social phenomenon (Trappmann et al., 2014).

A final aspect of sensitivity is known as social desirability. Respondents take the conformity of their own true answer with the applicable societal and social norms into account. Respondents then evaluate their own true answer as socially desirable or socially undesirable (Andersen & Mayerl, 2017; Tourangeau & Yan, 2007). In the case of a socially undesirable answer this can lead to respondents deviating from their true answer. This could lead to a response behaviour of respondents who tend to assess the social desirability of the potential answer options for an item in a questionnaire and accordingly choose the option categorised by them as the socially desirable answer. In this case, the responses of individual respondents would thus potentially provide information about the perception of social norms with regard to the opinion or behaviour asked about and could not serve as a benchmark for the actual spread of an opinion or behaviour in a population (Krumpal & Voss, 2020; Trappmann et al., 2014).

The three characteristics of sensitive survey topics mentioned above can lead to biased responses. Furthermore, such biases might occur systematically which can lead to incorrect conclusions regarding the prevalence of sensitive behaviour or to incorrectly assumed correlations between a sensitive trait and other social phenomena or demographic

characteristics (Tourangeau & Yan, 2007; Trappmann et al., 2014; Wolter, 2019; Wolter & Preisendörfer, 2013).

Surveys related to the COVID-19 pandemic in various countries showed consistently high compliance rates (Barari et al., 2020; Perrotta et al., 2020). These surveys mostly used direct questioning techniques for the self-assessment of respondents. Assuming that COVID-19-related compliance can be categorised as a sensitive survey topic, this approach may lead to systematic bias (Barari et al., 2020; Becher et al., 2020; Brouard et al., 2020; Daoust et al., 2020; Munzert & Selb, 2020).

Methodological features

Sensitivity research addresses the problem of biased data due to sensitive survey topics with special survey methods and particular formulations of items (Krumpal, 2013; Morin-Chassé et al., 2017; Näher & Krumpal, 2012; Trappmann et al., 2014; Wolter, 2019). In the following, one particular methodology known as the list experiment will be addressed because this has already been used to survey COVID-19-related compliance and could potentially be suitable for the data collection on corruption items.

In a list experiment, respondents are confronted with a list of questions or statements, for instance with a list of different behaviours. Respondents are asked to indicate only the number of behaviours that apply to them (in general or in a specific time interval) and not to evaluate each behaviour separately. Respondents are divided into two groups and given different lists. One group receives a shorter list of non-sensitive questions or statements (for comparison). The second group receives the same list which also contains the potentially sensitive item of interest. All respondents answer how many of the behaviours apply or applied to them. The extent of the sensitive behaviour can be calculated by comparing the mean values from the two groups. This approach is based on the assumption that the experimental group would have responded in the same way as the control group and that differences between the groups stem from the additional sensitive item (Blair & Imai, 2012; Glynn, 2013; Imai, 2011).

List experiments guarantee anonymity with regard to the individual answers of the respondents, unless respondents agree or disagree with all statements. When constructing a

list experiment, one should take care to ensure that one item from the short list is agreed to with a high probability and one item is disagreed to with a high probability. In this way it is possible to avoid respondents selecting zero or the maximum number of statements when asked about the number of applicable behaviours. In these two cases, the answer to all individual questions of the list experiment would be obvious. Otherwise, neither the interviewer (in the case of a face-to-face interview) nor the researchers who analyses the data know which of the behaviours were affirmed by the interviewee. This increase in anonymity - surpassing the confidentiality of responses in a survey - is also intended to encourage respondents to disclose socially undesirable behaviour (Glynn, 2013).

With the help of this technique, studies for various topics have been carried out, confirming some of the postulated change (Coutts & Jann, 2011; Kuha & Jackson, 2014; Rosenfeld et al., 2016; Wolter & Laier, 2014).

Selected list experiments

List experiments were also used for various surveys on COVID-19-related compliance and some results are presented below.¹⁵

Timmons et al. (2020a) used a list experiment in an online survey in Ireland with the aim of determining the social desirability of COVID-19-related compliance. The authors compared the results of direct questions on the following behaviours with the results of a list experiment: washing hands for at least 20 seconds or using hand sanitiser, keeping a minimum distance of two meters from people outside one's own household, wearing a mask in public, reducing personal social contacts (Timmons et al., 2020a, b). Significant differences were found between the survey methods. Among other things, Timmons et al. (2020b) found that according to the list experiment only 74.2 percent of respondents adhered to the contact restrictions, whereas 83.4 percent of respondents in the direct survey stated the same. With regard to the question on maintaining a minimum distance and hand hygiene, the results of the list experiments also showed lower compliance rates than in the direct survey. From these differences, the authors conclude the existence of social desirability in the form of an

¹⁵ At the beginning of this section, a special feature of the surveys during the pandemic should be pointed out. During the pandemic situation, it was not possible to conduct face-to-face interviews. The cited surveys were mainly online surveys. The transferability of the results to data collection in interview situations remains an empirical question that is not addressed in this paper.

overestimation of compliance. Furthermore, there was also a difference between the survey modes that did not correspond to the postulated direction. The authors found a difference of around 14 percent between the two survey methods with regard to wearing a mask. However, it was found that a lower proportion of respondents mentioned compliance with this regulation in the direct survey (54.2%) than in the list experiment (67.6%). The authors interpret this effect as a sign that respondents *did not* consider wearing a mask to be socially desirable at the time of their survey (Timmons et al., 2020b). Nevertheless, an effect of social desirability can also be observed for this item, as the prevalence differed between the respondents. The authors therefore conclude at the end of their article that a direct survey of all items would lead to biased or artificial compliance rates and that a list experiment is recommended (Timmons et al., 2020a).

Another example for the use of list experiments with regard to COVID-19-related compliance was presented by Larsen et al. (2020). The authors look at compliance with social distancing in the Danish population, using three different items. Larsen et al. (2020) also compare the results of two list experiments with respondents' answers to direct questions. In contrast to the first article by Timmons et al. (2020a), these comparisons show no differences between the two survey methods. The authors therefore conclude that there is no social desirability bias with regard to the selected behaviours and that direct questioning techniques therefore also lead to reliable findings on the extent of compliance.

The last example for the use of list experiments is the country comparing study by Becher et al. (2020). The authors collected data in nine countries (Australia, Great Britain, Germany, France, Italy, New Zealand, Austria, Sweden and the United States) using an online survey. The central question on social distancing was: *"I have met with two or more friends or acquaintances who do not live in my household"*. The authors compared the control and experimental groups of the list experiments per country and estimated the prevalence of country-specific compliance with the regulations in place at the time of the survey. In Germany, Sweden and Austria, the comparison resulted in a proportion of 42 percent to 64 percent of respondents who violated the regulation (in the case of Sweden violated the recommendation) of contact restriction. In Australia (33.6%), the United States (20.9%), France (12.5%) and New Zealand (12%), a minority of respondents stated that they had met friends or relatives, although the extent of this minority behaviour also differed significantly

between countries. In Italy and the UK, no significant prevalence of violating contact restrictions was found. The authors use their estimates as a starting point for further analyses regarding the relationship between compliance and other possible influencing factors. Becher et al. (2020) assume that the chosen approach was able to reduce biases due to socially desirable response behaviour and measurement errors.

As this exemplary list with regard to the use of list experiments shows, this technique was considered to be a useful tool for surveying COVID-19-related compliance by various authors and in some cases also led to a change in the estimate of the extent of the various behaviours. In the case of the Danish study the use of the list experiment did not lead to changes in prevalence but at the same time no negative effects of this questioning technique were found. In the other two studies, increased violations of various preventive regulations to combat the pandemic were found. Assuming that COVID-19-related compliance, as a sensitive topic, leads to biased and more socially desirable responses from respondents, this change can be considered a more realistic estimation.

In summary, the use of list experiments can therefore be seen as a useful complement that potentially leads to an improvement in the estimation of compliance.

Similarities in the investigation of corruption offences and compliance with COVID-19 regulations

Similarities can be identified in the survey of COVID-19-related and corruption-related compliance which subsequently also suggest the use of list experiments for the survey of corrupt behaviour. These similarities are presented in this chapter.

The focus on (past or general) behaviour of respondents – more precisely: compliance with anti-COVID-regulations on the one hand and compliance with anti-corruption norms or integrity regulations on the other – is an obvious similarity between the two topics.

Furthermore, the behaviours for both topics refer to social norms. On the one hand, these norms have grown over the long term and are firmly anchored in society, as in the case of the social ideal of not committing corrupt acts. On the other hand, the behavioural norms in relation to COVID-19 (hygiene regulations, social distancing) have only been in place for a

relatively short period of time. Nevertheless, the legislative regulations, which are (or can only be) partially monitored or enforced, lead to strong social expectations regarding the individual behaviour of each individual. As a result, this leads to a strong normative influence on both survey topics. Social desirability plays a central role when answering items on the topic of corruption as well as on compliance with various regulations to combat the COVID-19 pandemic. In both cases, a small(er) group benefits society to a certain extent if, for instance, a corrupt practice is used or a requirement to contain the pandemic is violated. This benefit (in the broadest sense) must be kept secret. With regard to compliance with anti-COVID-19 regulations, informal and formal sanctions are conceivable if respondents admit a violation in a survey. In the case of corruption, this circumstance is even exacerbated by possible legal sanctions for "confessing" to corrupt behaviour or bribery.

When conducting surveys, researchers often quite explicitly assure respondents that their answers will remain anonymous and that they will not be passed on to third parties. This is intended to reduce the threat of possible sanctions and increase the prospect of reliable answers. Survey research has already shown in various cases and with regard to various topics that these assurances are not always sufficient for respondents to actually admit to socially undesirable opinions or behaviour. The socially undesirable behaviour would have to be concealed with regard to both sets of questions discussed here. As a result, in the event of a violation of anti-COVID-19 regulations or anti-corruption regulations, a respondent may not want to admit this behaviour.

According to Tourangeau and Yan (2007), respondents' threat of disclosing their own behaviour (or opinions) can be categorised as sensitive content of questions. In this sense, questions about corrupt behaviour of respondents as well as COVID-19-related compliance can be classified as sensitive. The sensitivity of the corresponding question potentially depends on the respondent's answer (Tourangeau & Yan, 2007). In the case of corruption, the question on past corrupt behaviour is only sensitive for the proportion of respondents who have committed such corrupt acts. For the proportion of respondents who have not done so, however, a corresponding item in a questionnaire should be relatively unproblematic as long as they do not perceive the question as a social taboo. In the case of such a taboo, the topic in question would generally not be discussed openly which means that raising it during a survey could be perceived as inappropriate and offensive. In the case of such a topic,

respondents could also refuse to answer a question or avoid the question by editing responses. Although the two categories tend to merge into one another and the two aspects potentially mix with regard to a specific questionnaire item, it can be assumed for the topics discussed here that the threat of disclosure has a stronger impact on the answers of respondents.

Two preliminary conclusions can be drawn from this which apply to both corruption items and COVID-19-related items: On the one hand, the so-called social desirability bias can be assumed with regard to both sets of items. In studies on compliance with anti-COVID-19 regulations respondents reported high levels of compliance (Barari et al., 2020; Perrotta et al., 2020). Due to the social desirability of complying with the regulations, a proportion of these self-reports could possibly correspond to a socially desirable answer instead of a truthful statement about one's own behaviour. The same applies to corrupt behaviour and this response strategy is reinforced due to the threat of legal sanctions as well as the expectation of secrecy by accomplices. From the above, it can be deduced as a further similarity of the two survey topics that a substantial dark field can be assumed. The existence of this dark field indicates that the collected data may be a biased representation of reality. In the case of corruption and the violation of regulations to combat the pandemic, it can be assumed that the extent of this behaviour is underestimated for the reasons already mentioned.

Outlook

To summarise, compliance with preventive regulations to combat the COVID-19 pandemic as well as compliance in the sense of avoiding corruption were discussed as sensitive topics in the course of this article. The sensitivity of the two survey topics subsequently leads to potentially biased data in the case of a classic interview or questionnaire. This is due to the fact that respondents do not want to answer these items (at all) or do not want to answer them truthfully, as they (1) see them as a social taboo, (2) do not want to reveal their true answer due to a threat of disclosure or (3) are more likely to answer in a socially desirable manner than truthfully. These three aspects can lead to biased data in a classic survey. Nevertheless, researchers cannot identify per se from available data whether it is a (un)truthful answer from individual respondents. Potentially sensitive topics therefore always harbour the risk of biased data.

To minimise this risk, this article suggests using list experiments in surveys. These list experiments can increase the anonymity of respondents which has been shown to change the response behaviour of respondents in some studies on COVID-19-related compliance.

Due to the similarities of COVID-19-related and corruption-related compliance described above, list experiments are a worthwhile survey technique for corruption items as well. With list experiments, the probability of truthful answers from respondents can potentially be enhanced due to the increased anonymity. List experiments can therefore be an instrument for reducing the dark field in the area of corruption research and should therefore be considered for future research projects.

References

- Andersen, H., & Mayerl, J. (2017). Social desirability and undesirability effects on survey response latencies. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 135(1), 68-89.
- Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the covid- 19 epidemic? *The Lancet*, 395(10228), 931-934.
- Barari, S., Caria, S., Davola, A., Falco, P., Fetzer, T., Fiorin, S., Hensel, L., Ivchenko, A., Jachimowicz, J., King, G., Kraft-Todd, G., Ledda, A., MacLennan, M., Mutoi, L., Pagani, C., Reutskaya, E., & Slepoi, F. R. (2020). Evaluating covid-19 public health messaging in Italy: Self-reported compliance and growing mental health concerns. *medRxiv*. doi: 10.1101/2020.03.27.20042820.
- Becher, M., Stegmueller, D., Brouard, S., & Kerrouche, E. (2020). Comparative experimental evidence on compliance with social distancing during the COVID-19 pandemic. *medRxiv* doi: 10.1101/2020.07.29.20164806).
- Beck, L., & Nagel, V. (2019). Korruption aus ökonomischer Perspektive. In P. Graeff & T. Rabl (Hrsg.), *Was ist Korruption? Begriffe, Grundlagen und Perspektiven gesellschaftswissenschaftlicher Korruptionsforschung* (S. 39–58). Nomos.

- Betsch, C. (2020). How behavioural science data helps mitigate the COVID19 crisis. *Nature Human Behaviour*, 4, 438.
- Blair, G., & Imai, K. (2012). Statistical analysis of list experiments. *Political Analysis*, 20(1), 47-77.
- Böckenholt, U. (2014). Modelling motivated misreports to sensitive survey questions. *Psychometrika*, 79(3), 515-537.
- Brouard, S., Vasilopoulos, P., & Becher, M. (2020). Sociodemographic and psychological correlates of compliance with the covid-19 public health measures in France. *Canadian Journal of Political Science*, 53(2), 253- 258.
- Coutts, E., & Jann, B. (2011). Sensitive questions in online surveys: Experimental results for the randomised response technique (RRT) and the unmatched count technique (UCT). *Sociological Methods and Research*, 40(1), 169-193.
- Daoust, J.-F., Nadeau, R., Dassonneville, R., Lachapelle, E., Bélanger, E., Savoie, J., & van der Linden, C. (2020). How to survey citizens' compliance with COVID-19 public health measures? evidence from three survey experiments. *Journal of Experimental Political Science*. doi:10.1017/XPS.2020.25.
- Glynn, A. N. (2013). What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opinion Quarterly* 77, 159-172.
- Graeber, D., Schmidt-Petri, C., & Schröder, C. (2020a). Hohe Impfbereitschaft gegen Covid-19 in Deutschland, Impfpflicht bleibt kontrovers, *SOEP papers on Multidisciplinary Panel Data Research*, No. 1103. Deutsches Institut für Wirtschaftsforschung (DIW).
- Graeber, D., Schmidt-Petri, C., & Schröder, C. (2020b). Covid-19: Mehrheit der Deutschen würde sich freiwillig impfen lassen, die Hälfte ist für eine Impfpflicht. *DIW aktuell*, No. 54. Deutsches Institut für Wirtschaftsforschung (DIW).
- Graeff, P., & Dombois, R. (2019). Soziologische Zugänge zur Korruptionsproblematik. In P. Graeff & T. Rabl (Hrsg.), *Was ist Korruption? Begriffe, Grundlagen und Perspektiven gesellschaftswissenschaftlicher Korruptionsforschung* (S. 123–142). Nomos.

- Graeff, P., & Steßl, A. (2017). Effektive Compliance: Ursachen, Hindernisse und Lösungsvorschläge. In C. Stark (Hrsg.), *Korruptionsprävention: Klassische und ganzheitliche Ansätze*. Springer Fachmedien.
- Hipp, L., Bünning, M., Munnes, S., & Sauermann, A. (2020). Problems and pitfalls of retrospective survey questions in COVID-19 studies. *Survey Research Methods*, 14(2), 109-114.
- Huebner, M., Spieß, C. K., Siegel, N. A., & Wagner, G. G. (2020). Wohlbefinden von Familien in Zeiten von Corona: Eltern mit jungen Kindern am stärksten beeinträchtigt. *DIW Wochenbericht*, No. 30+31. Deutsches Institut für Wirtschaftsforschung (DIW).
- Imai, K. (2011). Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association*, 106(494), 407-416.
- Jann, B., Krumpal, I., & Wolter, F. (2019). Editorial: Social desirability bias in surveys collecting and analysing sensitive data. *methods, data, analyses*, 13(1), 3-6.
- Juhl, S., Lehrer, R., Blom, A. G., Wenz, A., Rettig, T., Reifenscheid, M., Naumann, E., Möhring, K., Krieger, U., Friedel, S., Fikel, M., & Cornesse, C. (2020). Die Mannheimer Corona-Studie: Gesellschaftliche Akzeptanz politischer Maßnahmen und befürchtete Konsequenzen für die Wirtschaft. https://madoc.bib.uni-mannheim.de/55140/1/2020-0403_Wochenbericht.pdf. Accessed: 15 Apr. 2021.
- Kapteyn, A., Angrisani, M., Bennett, D., Bruine de Bruin, W., Darling, J., Gutsche, T., Liu, Y., Meijer, E., Perez-Arce, F., Schaner, S., Thomas, K., & Weerman, B. (2020). Tracking the effect of the COVID-19 pandemic on the lives of American households. *Survey Research Methods*, 14(2), 179-186.
- Kühne, S., Kroh, M., Liebig, S., & Zinn, S. (2020). The need for household panel surveys in times of crisis: The case of SOEP-CoV. *Survey Research Methods*, 14(2), 195-203.
- Kuha, J., & Jackson, J. (2014). The item count method for sensitive survey questions: Modelling criminal behaviour. *Journal of the Royal Statistical Society*, 63(2), 321-341.
- Kreuter, F., Barkay, N., Bilinski, A., Bradford, A., Chiu, S., Eliat, R., Fan, J., Galili, T., Haimovich, D., Kim, B., LaRocca, S., Li, Y., Morris, K., Presser, S., Sarig, T., Salomon, J. A., Stewart,

- K., Stuart, E. A., & Tibshirani, R. (2020). Partnering with Facebook on a university-based rapid turn-around global survey. *Survey Research Methods*, 14(2), 159-163.
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys. A literature review. *Quality & Quantity*, 47(4), 2025-2047.
- Krumpal, I., & Voss, T. (2020). Sensitive questions and trust: Explaining respondents' behaviour in randomized response surveys. *SAGE Open*. doi: 10.1177/2158244020936223.
- Larsen, M., Nyrup, J., & Petersen, M. B. (2020). Do survey estimates of the public's compliance with COVID-19 Regulations suffer from social desirability bias? *Journal of Behavioural Public Administration*, 3(2), 1-9.
- Morin-Chassé, A., Bol, D., Stephenson, L. B., & St-Vincent, S. L. (2017). How to survey about electoral turnout? The efficacy of the face-saving response items in 19 different contexts. *Political Science Research and Methods*, 5(3), 575-584.
- Munzert, S., & Selb, P. (2020). Can we directly survey adherence to non-pharmaceutical interventions? Evidence from a list experiment conducted in Germany during the early corona pandemic. *Survey Research Methods*, 14(2), 205-209.
- Näher, A.-F., & Krumpal, I. (2012). Asking sensitive questions: The impact of forgiving wording and question context on social desirability bias. *Quality & Quantity*, 46(5), 1601-1616.
- Naumann, E., Mata, J., Reifenscheid, M., Möhring, K., Wenz, A., Rettig, T., Lehrer, R., Krieger, U., Juhl, S., Friedel, S., Fikel, M., Cornesse, C., & Blom, A. G. (2020). Die Mannheimer Corona-Studie: Schwerpunktbericht zum Angstempfinden in der Bevölkerung. http://www.uni-mannheim.de/media/Einrichtungen/gip/Corona_Studie/Schwerpunktbericht_Angstempfinden_Mannheimer_Corona_Studie.pdf. Accessed: 7 Apr. 2021.
- Neumann-Böhme, S., Varghese, N. E., Sabat, I., Barros, P. P., Brouwer, W., van Exel, J., Schreyögg, J., & Stargardt, T. (2020). Once we have it, will we use it? A European survey on willingness to be vaccinated against COVID-19. *The European journal of health economics*, 21(7), 977-982.

- Perrotta, D., Grow, A., Rampazzo, F., Cimentada, J., Del Fava, E., GilClavel, S. & Zagheni, E. (2020). Behaviours and attitudes in response to the covid-19 pandemic: Insights from a cross-national facebook survey. *medRxiv*. doi: 10.1101/2020.05.09.20096388.
- Post, J. C., Class, F., & Kohler, U. (2020). Unit nonresponse biases in estimates of SARS-CoV-2 prevalence. *Survey Research Methods*, 14(2), 115-121.
- Richards, L. (2017). Using Survey Methods to Research Corruption. In A. K. Schwickerath, A. Varraich & L.-E. Smith (Eds.), *How to research corruption? Conference Proceedings: Interdisciplinary Corruption Research Forum*, June 2016, 4-12.
- Rosenfeld, B., Imai, K., & Shapiro, J. (2016). An empirical validation study of popular survey methodologies for sensitive questions. *American Journal of Political Science*, 60(3), 783-802.
- Sibley, C. G., Greaves, L. M., Satherley, N., Wilson, M. S., Overall, N. C., Lee, C. H. J., Milojeć, P., Bulbulia, J., Osborne, D., Milfont, T. L., Houkamau, C. A., Duck, I. M., Vickers-Jones, R., & Barlow, F. K. (2020). Effects of the COVID-19 pandemic and nationwide lockdown on trust, attitudes toward government, and well-being. *American Psychologist*, 75(5), 618-630.
- Skarbek-Kozietulska, A., Preisendörfer, P., & Wolter, F. (2012). Leugnen oder gestehen? Bestimmungsfaktoren wahrer Antworten in Befragungen. *Zeitschrift für Soziologie*, 41(1), 5–23.
- Timmons, S., McGinnity, F., Belton, C., Barjaková, M., & Lunn, P. (2020a). It depends on how you ask: Measuring bias in population surveys of compliance with COVID-19 public health guidance. *Journal of Epidemiology and Community Health*, 75, 387-389.
- Timmons, S., McGinnity, F., Belton, C., Barjaková, M., & Lunn, P. (2020b). It depends on how you ask: Measuring bias in population surveys of compliance with COVID-19 public health guidance. *Supplementary Material*. doi:10.1136/jech-2020-215256
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859-883.

- Trappmann, M., Krumpal, I., Kirchner, A., & Jann, B. (2014). Item sum. A new technique for asking quantitative sensitive questions. *Journal of Survey Statistics and Methodology*, 2(1), 58-77.
- Utych, S. M., & Fowler, L. (2020). Age-based messaging strategies for communication about COVID-19. *Journal of Behavioural Public Administration*, 3(1), 1-14.
- Wolter, F. (2019). A New Version of the Item Count Technique for Asking Sensitive Questions Testing the Performance of the Person Count Technique. *methods, data, analyses*, 13(1), 169-192.
- Wolter, F., & Laier, B. (2014). The effectiveness of the item count technique in eliciting valid answers to sensitive questions. An evaluation in the context of self-reported delinquency. *Survey Research Methods*, 8(3), 153-168.
- Wolter, F., & Preisendörfer, P. (2013). Asking sensitive questions: An evaluation of the randomised response technique versus direct questioning using individual validation data. *Sociological Methods & Research*, 42(3), 321-353.

2.2 German Original

Original publication: Dunkelmann, F. (2022). Korruption und Corona. Mögliche Gemeinsamkeiten zweier Befragungsthemen. In: Wolf, S., Graeff, P. (Hrsg.) Corona und Korruption. Springer VS, Wiesbaden. https://doi.org/10.1007/978-3-658-35664-4_7.

Korruption und Corona. Mögliche Gemeinsamkeiten zweier Befragungsthemen

Einleitung

„Corruption is by no means a mainstream topic in national surveys...”

(Richards, 2017)

Die Erforschung und Aufdeckung von Korruption stellt Forschende seit jeher vor verschiedene Herausforderungen. Die bestehenden Hürden für Korruptionsforschung und -aufdeckung, beispielsweise die fehlende Kenntnis zum Ausmaß von Korruption in einer Gesellschaft oder in einem Unternehmen, werden durch die Corona-Pandemie zum Teil verstärkt. Gleichzeitig lassen sich aus Sicht der Umfrageforschung auch wichtige Lehren für die Erhebung von Korruptionsdaten aus Erhebungen während der Pandemie ableiten. Erhebungen mit Bezug zur Corona-Pandemie nutzen zum Teil Fragetechniken, die auch für eine Datensammlung in Umfragen zu korrupten Verhaltensweisen aussichtsreich und lohnenswert sein könnten. Diese Potenziale sollen im Rahmen dieses Beitrags aus der Gegenüberstellung von Verhaltensweisen, die gegen Compliance in beiden Bereichen verstoßen, abgeleitet werden.

Unter Compliance mit Bezug zur Korruption soll dabei, ähnlich dem Begriffsverständnis in anderen Themengebieten, die Einhaltung und Befolgung von Regeln, Maßnahmen sowie Standards verstanden werden (Graeff & Steßl, 2017). Diese Maßnahmen richten sich dabei gegen Korruption in ihren verschiedenen Facetten. Unter dem Begriff Korruption kann in diesem Zusammenhang ein Verhalten von Akteuren, genauer ein kooperativer sowie

devianter Akt zwischen mindestens zwei Individuen, verstanden werden. Die Kooperation nimmt den Charakter devianten Verhaltens an, wenn Beteiligte als Vertreter eines bestimmten Amtes handeln und gleichzeitig von den Rollenverpflichtungen ebendieses Amtes abweichen (Graeff & Dombois, 2019). Die Kooperation kann sich zudem zum Nachteil unbeteiligter Dritter auswirken (Beck & Nagel, 2019). Prominente Beispiele für Akte der Korruption sind dabei etwa die Bestechung von politischen Amtsträgern, das Umgehen öffentlicher Ausschreiben oder das Austauschen von gegenseitigen Gefallen im beruflichen Kontext. Im Kern ist korruptionsbezogene Compliance somit als Verhalten von Akteuren zu verstehen, die sich (aktiv) einer sozialen Norm oder einem Gesetz folgend gegen die Beteiligung an korrupten Verhaltensweisen entscheiden.

Eine Datenerhebung zum Ausmaß von praktizierter Compliance oder korrupter Verhaltensweisen in einem Unternehmen oder in einem Land gestaltet sich aus verschiedenen Gründen schwierig. Die direkte Abfrage korrupter Verhaltensweisen ist problematisch, da Korruption die Verschwiegenheit gegenüber unbeteiligten Dritten impliziert sowie eine sozial unerwünschte Handlung darstellt. Dies führt dazu, dass die Erhebung in Form von konkreten Fragen nach korrupten Verhaltensweisen von Befragten eher selten stattfindet. Häufiger sind hingegen Items zur Beurteilung des Ausmaßes von Korruption, die sich nicht direkt nach dem Verhalten der Befragten erkundigen (Richards, 2017). Stattdessen werden Befragte beispielsweise um eine Einschätzung des Ausmaßes der Korruption innerhalb ihres Landes gebeten. Im Rahmen dieses Beitrags wird eine Gelegenheit aufgezeigt, wie trotz bestehender Hürden auch konkrete Fragen nach korrupten Verhaltensweisen von Befragten möglich werden.

Diese Möglichkeit leitet sich aus dem Vorgehen bei der Erhebung von Compliance in Bezug auf COVID-19 ab. Hinsichtlich der Corona-Pandemie ist mit dem Schlagwort der Compliance die Einhaltung präventiver, nicht-medizinischer Maßnahmen zur Eindämmung des Infektionsgeschehens (z. B. regelmäßiges Händewaschen, Tragen von Mund-Nasen Bedeckungen, Abstandsgebot, Kontaktbeschränkungen) gemeint. Diese Regelbefolgung wird als wichtiger Baustein der Pandemiebekämpfung betrachtet (Anderson et al., 2020; Betsch, 2020). Eine objektive Messung hinsichtlich der Einhaltung dieser Maßnahmen ist jedoch schwierig, da sich Maßnahmen beispielsweise auch auf den privaten Raum der Bürgerinnen und Bürger beziehen (Timmons et al., 2020a). Larsen et al. (2020) führen beispielsweise aus,

dass mit Hilfe von Handydaten einerseits die Bewegungsmuster der Bevölkerung abgebildet und somit die Einhaltung von Ausgangssperren oder anderer Maßnahmen geschätzt werden kann. Andererseits ermöglichen diese Daten keine Auskunft über das konkrete Verhalten von Individuen, wenn diese aufeinandertreffen. Es kann aus den Bewegungsdaten nicht geschlussfolgert werden, ob Individuen sich umarmen, sich die Hand zur Begrüßung geben oder konsequent Masken bei Kontakten außerhalb des eigenen Haushaltes tragen. Weiterhin kann auch das Verhalten von Individuen in ihren eigenen Haushalten, wie beispielsweise das Waschen der Hände bei der Rückkehr in den eigenen Haushalt, anhand dieser Daten nicht erhoben werden. Um Informationen über diese Art der Compliance zu bekommen, sind Erhebungsdaten im Rahmen von Bevölkerungsumfragen nötig.

Aus diesem Grund sind derartige Fragen auch zum Bestandteil diverser Erhebungsinstrumente geworden, die eine Datengrundlage bezüglich des Ausmaßes verschiedener Verhaltensweisen liefern sollen. In diesem Zusammenhang wurden auch spezifische Befragungstechniken aus dem Bereich der Sensitivitätsforschung verwendet, die im Laufes dieses Beitrags vorgestellt werden (Becher et al., 2020; Larsen et al., 2020; Timmons et al., 2020b). Anschließend wird anhand von Gemeinsamkeiten zwischen Korruptionsitems und Surveyfragen mit Coronabezug hergeleitet, welche Potenziale für die Erhebung korrupter Verhaltensweisen, Intentionen und Tendenzen abgeleitet werden können.

Survey Research zur Corona-Pandemie

Thematische Vielfalt

Seit dem Ausbruch der Corona-Pandemie im Frühling 2020 wurden zahlreiche Umfragen mit Bezügen zu dieser Ausnahmesituation durchgeführt. Die Fragen in diesen Erhebungen decken dabei ein breites Spektrum ab und sollen an dieser Stelle zunächst überblicksartig betrachtet werden. Es lassen sich insgesamt 6 Kategorien identifizieren:

- 1 Fragen zu verschiedenen, allgemeinen Befragungsthemen, die bereits vor der Corona-Pandemie erhoben wurden;
- 2 Fragen zur Infektion mit COVID-19;
- 3 Fragen zu (potenziellen) Auswirkungen der Corona-Pandemie;

- 4 Fragen zur Meinung hinsichtlich (politischer) Maßnahmen der Pandemiebekämpfung (exklusive medizinischer Maßnahmen);
- 5 Fragen zur Einhaltung von (politischen) Maßnahmen der Pandemiebekämpfung (exklusive medizinischer Maßnahmen);
- 6 Fragen zu (potenziellen) Impfstoffen und zur Impfbereitschaft.

Fragen der ersten Kategorie sind deckungsgleich mit Items, die bereits vor der Pandemie erhoben wurden. Diese Items werden hinsichtlich potenzieller Veränderungen betrachtet, die auf die Pandemiesituation, konkrete politische Entscheidungen, Kontaktbeschränkungen oder (verschiedene Stadien von) Lockdowns zurückgeführt werden. Beispielhaft rücken Sibley et al. (2020) die Auswirkungen der Pandemie auf die Lebenszufriedenheit, die Gesundheit sowie das institutionelle Vertrauen von Befragten in den Mittelpunkt der Betrachtung. Perrotta et al. (2020) betrachten weiterhin das Vertrauen in verschiedene Institutionen sowie das Ausmaß der Verlässlichkeit verschiedener Informationsquellen (wie z. B. Zeitungen, soziale Medien, Rundfunk, Erzählungen von Freunden oder Bekannten). Zudem werden regelmäßige Studien, wie beispielsweise das Sozio-ökonomische Panel (SOEP) fortgeführt, um Veränderungen diverser Variablen festzuhalten (Kühne et al., 2020).

Fragen der zweiten Kategorie beziehen sich hingegen auf (aktuelle sowie vergangene) Infektionen mit COVID19, potenzielle Krankheitsverläufe im Falle einer Infektion von Befragten sowie das Infektionsgeschehen im persönlichen Umfeld (Kapteyn et al., 2020; Kreuter et al., 2020). Die Erhebung von Umfragen zu diesem Themenkomplex rücken beispielsweise Kreuter et al. (2020) in den Mittelpunkt und begründen dies mit einer schnelleren Erhebung im Vergleich zu offiziellen Statistiken. Die Autor*innen entwickelten ein Befragungsinstrument, welches über Facebook täglich Befragte in mehr als 200 Territorien erreichte. Weiterhin konnte aufgrund der Zusammenarbeit mit Facebook eine spezifische Gewichtung der Daten vorgenommen werden. Auf diese Weise erheben die Autor*innen Daten, die zum Teil genauere Aussagen über das aktuelle Infektionsgeschehen ermöglichen, als dies offizielle Statistiken, sofern in einem Land derartige (staatliche sowie technische) Strukturen vorhanden sind, leisten können. Weiterhin setzen sich Post et al. (2020) mit möglichen Fehlerquellen bei der Erhebung offizieller Statistiken auseinander und bekräftigen im Anschluss die Erhebung von Umfragedaten zur verbesserten Abbildung des Infektionsgeschehens.

Fragen der dritten Kategorie fokussieren auf verschiedene Auswirkungen der Corona-Pandemie sowie der entsprechend geltenden Maßnahmen der Pandemiebekämpfung. Items

dieser Kategorie fragen dabei einerseits nach der Einschätzung der/ des Befragten hinsichtlich zu erwartender Auswirkungen in der Zukunft, beispielsweise hinsichtlich wirtschaftlicher oder persönlicher Folgen (Juhl et al., 2020). Andererseits beziehen sich Items dieser Kategorie auch auf Veränderungen des alltäglichen Lebens zum Zeitpunkt der Befragung. Diese Items zielen dabei beispielhaft auf potenzielle Veränderungen des Familienlebens (aufgrund von Kindertagesstätten- oder Schulschließungen) oder des Arbeitslebens (aufgrund von Kontaktbeschränkungen, Homeoffice, Kurzarbeit etc.) ab (Hipp et al., 2020; Huebner et al., 2020; Kapteyn et al., 2020; Kühne et al., 2020). Weiterhin werden Fragen zur emotionalen, mentalen und psychischen Verfassung von Befragten eingesetzt (Hipp et al., 2020; Kreuter et al., 2020; Kühne et al., 2020; Naumann et al., 2020; Perrotta et al., 2020).

Weitere Fragen, die als vierte Kategorie zusammengefasst werden, beziehen sich auf das Verständnis, die Akzeptanz und die Zustimmung der jeweils zum Zeitpunkt der Befragung geltenden (politischen) Maßnahmen der Pandemiebekämpfung. Bei dieser Kategorie handelt es sich zunächst um die Erstellung eines allgemeinen Meinungsbildes hinsichtlich getroffener Entscheidungen sowie deren Kommunikation an die Bürgerinnen und Bürger. Antworten zu dieser Fragenkategorie sind somit auch aus Sicht der politischen Entscheider*innen von großer Bedeutung und werden beispielsweise in wöchentlichen Befragungen im Rahmen der Mannheimer Corona-Studie erhoben (Juhl et al., 2020).

Fragen der fünften und im Rahmen dieses Beitrags zentralen Kategorie fokussieren die *Einhaltung* verschiedener (politischer) Maßnahmen zur Eindämmung der Pandemie. Diese Fragen gehen somit über die *Meinung* der Befragten hinaus und beziehen sich detaillierter auf vergangene oder zukünftige Verhaltensweisen. Beispielhaft werden Fragen nach Kontakten zu Freunden oder Verwandten trotz eines Gebots zur Kontaktbeschränkung (Daoust et al., 2020; Kreuter et al., 2020; Munzert & Selb, 2020; Perrotta et al., 2020) oder Fragen hinsichtlich der Handhygiene von Befragten (Kapteyn et al., 2020; Utych & Fowler, 2020) dieser Kategorie zugeordnet. Weiterhin können auch Fragen nach dem Tragen von Mund-Nasen-Bedeckungen, die ebenfalls als präventives Mittel kategorisiert werden, dieser Fragengruppen zugeordnet werden (Kapteyn et al., 2020). Die erfragten Verhaltensweisen, welche im Englischen als „*nonpharmaceutical interventions*“ bezeichnet werden, gelten der Prävention und waren zu Beginn der Pandemie die einzige Möglichkeit zu deren Eindämmung (Anderson et al., 2020; Betsch, 2020).

In Abgrenzung dazu ist die Pandemiebekämpfung mittels einer Impfung als zweiter Baustein von zentraler Bedeutung für das Krisenmanagement. Auf ebendiese Möglichkeit beziehen sich die Fragen der sechsten Kategorie. Dieser Kategorie werden sowohl verschiedene Fragen zu Impfstoffen als auch Items zur Impfbereitschaft von Befragten zugeordnet (Graeber et al., 2020a, b; Kapteyn et al., 2020; Neumann-Böhme et al., 2020).

Sensitivität

Die vorangestellte Auflistung verdeutlicht die Bandbreite verwendeter Befragungsthemen sowie -themen. Aus Sicht der Umfrageforschung sind ausgewählte Kategorien besonders interessant, da in einigen Bereichen von den Pfaden der klassischen Umfragetechniken abgewichen wird. Im Folgenden soll vor allem auf die Abfrage hinsichtlich der *Einhaltung* von Maßnahmen zur Minimierung von Ansteckungsgefahren mit COVID-19 eingegangen werden. Diese erfragten Verhaltensweisen, wie beispielsweise Handhygiene oder Kontaktbeschränkungen, sind dabei von zentraler Bedeutung für die Entwicklung der Pandemie – sowohl solange keine Impfstoffe verfügbar waren, als auch im Verlauf von Impfkampagnen – und zeichnen sich aus Sicht der Umfrageforschung durch besondere Charakteristika aus.

Eines dieser Charakteristika ist die Sensitivität der Thematiken, bei denen es um Compliance geht. Die Sensitivität von Items führt zu (potenziell) verzerrten Daten, wenn befragte Personen auf diese Inhalte nicht oder nicht wahrheitsgemäß antworten. In Anlehnung an Tourangeau und Yan (2007) lassen sich drei Dimensionen von Sensitivität unterscheiden.

Zunächst können Items von Befragten als aufdringlich und der Befragungssituation unangemessen wahrgenommen werden. Das trifft insbesondere zu, wenn Fragen sich auf gesellschaftliche Tabus beziehen. Befragte können den Inhalt der Frage dann als gesellschaftliches Tabu, über welches man allgemein nicht sprechen möchte, ansehen. In der Konsequenz verweigern Befragte zum Teil die Aussage, wodurch erhobene Daten verzerrt sein können (Jann et al., 2019; Tourangeau & Yan, 2007).

Weiterhin kann die Sensitivität von Items mit einer Angst vor Entdeckung beziehungsweise Bekanntwerden von gegebenen Antworten verbunden sein. Einige Antworten auf spezifische Fragen, die etwa strafrechtliche Relevanz besitzen (und für die man auch lange nach der Tat

belangt werden könnte), können aus Sicht von Befragten daher heikel erscheinen. Der Grad der Sensitivität einer Frage variiert zwischen einzelnen Befragten und hängt von deren potenzieller Antwort ab (Tourangeau & Yan, 2007). Befragte, deren wahre Antwort in diesem Sinne sensitiv ist, können ebenfalls eine Antwort verweigern. Möglicherweise kann eine Antwortverweigerung auf eine Frage, bei der nur eine der Antwortmöglichkeiten als sensitiv eingeschätzt wird, als eine Art Schuldeingeständnis oder Signal für eine bestimmte wahre Antwort betrachtet werden (Tourangeau & Yan, 2007; Wolter, 2019). Daher liegt es nahe, dass Befragte in diesem Fall eine Antwortmöglichkeit wählen, die nicht der Wahrheit entspricht (Böckenholt, 2014). Beispielsweise könnten Befragte, die eine Straftat begangen haben, ebendiese in einer Umfrage verschweigen (Skarbek-Kozietulska et al., 2012). Auch diese Dimension der Sensitivität führt somit zu potenziell verzerrten Daten, wobei die Verzerrung über die Item-Nonresponse hinausgeht und auch die gegebenen Antworten ein fehlerhaftes Bild eines sozialen Phänomens zeichnen können (Trappmann et al., 2014).

Ein letzter Aspekt der Sensitivität ist unter dem Begriff der sozialen Erwünschtheit bekannt. Befragte beziehen dabei die Übereinstimmung der eigenen, wahren Antwort mit den geltenden gesellschaftlichen und sozialen Normen ein. Befragte beurteilen dann ihre eigene, wahre Antwort als sozial erwünscht oder sozial unerwünscht (Andersen & Mayerl, 2017; Tourangeau & Yan, 2007). Das kann im Falle einer sozial unerwünschten Antwort dazu führen, dass Befragte von ihrer wahren Antwort abweichen. Dies könnte zu einem Antwortverhalten von Befragten führen, die bei einem Item in einem Fragebogen eher die soziale Erwünschtheit der potenziellen Antwortmöglichkeiten einschätzen und entsprechend die von ihnen als sozial erwünschte Antwort kategorisierte Möglichkeit wählen. In diesem Fall wären die Antworten von einzelnen Befragten somit potenziell eine Auskunft über die Wahrnehmung der sozialen Normen hinsichtlich der erfragten Meinung oder Verhaltensweisen und könnten nicht als Maßstab für die tatsächliche Verteilung einer Meinung oder Verhaltensweise in einer Grundgesamtheit dienen (Krumpal & Voss, 2020; Trappmann et al., 2014).

Die drei genannten Charakteristika sensitiver Befragungsinhalte führen zu potenziell verzerrten Antworten von Befragten. Weiterhin können derartige Verzerrungen, die möglicherweise systematisch auftreten, zu falschen Schlussfolgerungen hinsichtlich des Ausmaßes von sensitiven Verhaltensweisen oder zu fälschlicherweise angenommenen

Zusammenhängen zwischen sensitiven Verhaltensweisen und anderen sozialen Phänomenen oder demografischen Merkmalen führen (Tourangeau & Yan, 2007; Trappmann et al., 2014; Wolter, 2019; Wolter & Preisendörfer, 2013).

Umfragen mit Bezug zur Corona-Pandemie in verschiedenen Ländern zeigten einheitlich hohe Compliance-Raten (Barari et al., 2020; Perrotta et al., 2020). In diesen Umfragen wurden zumeist direkte Fragetechniken für die Selbsteinschätzung der Befragten verwendet. Unter der Annahme, dass coronabezogene Compliance als sensitives Befragungsthema kategorisiert werden kann, führt dieses Vorgehen möglicherweise zu einer systematischen Verzerrung (Barari et al., 2020; Becher et al., 2020; Brouard et al., 2020; Daoust et al., 2020; Munzert & Selb, 2020).

Methodische Besonderheiten

Der Problematik von verzerrten Daten aufgrund von sensitiven Befragungsinhalten begegnet die Sensitivitätsforschung mit speziellen Befragungsmethoden sowie spezifischen Formulierungen von Items (Krumpal, 2013; Morin-Chassé et al., 2017; Näher & Krumpal, 2012; Trappmann et al., 2014; Wolter, 2019). Im Folgenden soll vor allem auf eine Methodik, das sogenannte List Experiment eingegangen werden, da dieses bereits für die Erhebung coronabezogener Compliance genutzt wurde und sich potenziell für die Erhebung von Korruptionselementen ebenfalls eignen könnte.

In einem List Experiment wird Befragten eine Liste von Fragen oder Statements, etwa eine Reihe verschiedener Verhaltensweisen, vorgelegt. Im Anschluss werden die Befragten beispielsweise gebeten ausschließlich die Anzahl, der auf sie (allgemein oder in einem bestimmten Zeitintervall) zutreffenden Verhaltensweisen, anzugeben und nicht für jede genannte Verhaltensweise eine Aussage zu treffen. Die Befragten werden in zwei Gruppen eingeteilt, wobei sich die vorgelegten Listen unterscheiden. Eine Gruppe erhält eine kürzere Liste mit nichtsensitiven Fragen oder Statements, die dem späteren Vergleich dient. Die zweite Gruppe erhält ebendiese Liste, die zusätzlich das interessierende und potenziell sensitive Item enthält. Alle Befragten beantworten, wie viele der Verhaltensweisen auf sie zutreffen beziehungsweise zutrafen. Mittels eines Vergleichs der Mittelwerte aus den beiden Gruppen kann das Ausmaß der sensitiven Verhaltensweise eingeschätzt werden. Diese

Vorgehensweise unterliegt der Annahme, dass die Experimentalgruppe in gleicher Weise wie die Kontrollgruppe geantwortet hätte und Unterschiede zwischen den Gruppen auf das zusätzliche, sensitive Item zurückzuführen sind (Blair & Imai, 2012; Glynn, 2013; Imai, 2011).

List Experimente gewährleisten die Anonymität hinsichtlich der individuellen Antworten der Befragten, sofern Befragte nicht allen Statements zustimmen oder diese ablehnen. Bei der Konstruktion eines List Experimentes sollte darauf geachtet werden, dass jeweils ein Item der kürzeren Liste mit hoher Wahrscheinlichkeit bejaht sowie eines ebenfalls mit hoher Wahrscheinlichkeit verneint wird. Auf diese Weise kann vermieden werden, dass Befragte beider Gruppen auf die Frage nach der Anzahl der zutreffenden Verhaltensweisen mit null oder mit der maximalen Anzahl antworten. In diesen beiden Fällen wäre die Antwort auf das List Experiment eindeutig auf alle Einzelfragen zurück zu beziehen. Andernfalls wissen weder Interviewer (im Fall eines Face-to-Face Interviews) noch Forschende, die die Daten später auswerten, welche der Verhaltensweisen von der befragten Person bejaht wurden. Mittels dieser Steigerung der Anonymität – über die Vertraulichkeit von Antworten in einer Umfrage hinaus – sollen Befragte auch zu einer Offenlegung sozial unerwünschter Verhaltensweisen bewegt werden (Glynn, 2013).

Mithilfe dieser Technik wurden in der Vergangenheit für verschiedene Themen Untersuchungen durchgeführt, die diese postulierte Veränderung zum Teil bestätigten (Coutts & Jann, 2011; Kuha & Jackson, 2014; Rosenfeld et al., 2016; Wolter & Laier, 2014).

Ausgewählte List Experimente

List Experimente wurden ebenfalls für verschiedene Erhebungen zur coronabezogenen Compliance genutzt, deren Ergebnisse im Folgenden beispielhaft dargestellt werden sollen.¹⁶ Timmons et al. (2020a) setzten ein List Experiment bei einer Onlinebefragung in Irland ein mit dem Ziel, die soziale Erwünschtheit coronabezogener Compliance zu ermitteln. Die Autor*innen verglichen dabei die Ergebnisse direkter Fragen zu folgenden Verhaltensweisen mit den Resultaten eines List Experimentes: Hände waschen für mindestens 20s

¹⁶ Zu Beginn dieses Abschnitts sei auf eine Besonderheit der Umfragen zu Zeiten der Pandemie hingewiesen. Während der Pandemiesituation bestand keine Möglichkeit zu Face-to-Face-Interviews. Die zitierten Umfragen fanden schwerpunktmäßig als Onlinebefragungen statt. Die Übertragbarkeit der Ergebnisse auf Datenerhebungen in Interviewsituationen bleibt eine empirisch zu überprüfende Frage, die im Rahmen dieser Arbeit nicht bearbeitet wird.

beziehungsweise die Nutzung von Handdesinfektionsmitteln, die Einhaltung eines Mindestabstandes von zwei Metern zu Personen außerhalb des eigenen Haushaltes, das Tragen einer Maske in der Öffentlichkeit, die Reduzierung der persönlichen, sozialen Kontakte (Timmons et al., 2020a, b). Für die Verhaltensweisen wurden dabei signifikante Unterschiede zwischen den Befragungsmethoden gefunden. Unter anderem fanden Timmons et al. (2020b) heraus, dass sich die Befragten im List Experiment nur zu 74,2% an die Kontaktbeschränkungen hielten, wohingegen in der direkten Abfrage 83,4% der Befragten ebendies angaben. Bezüglich der Frage zur Einhaltung eines Mindestabstandes sowie zur Handhygiene zeigten sich in den Ergebnissen der List Experimente ebenfalls geringere Complianceraten als in der direkten Befragung. Aus diesen Differenzen schließen die Autor*innen das Vorhandensein sozialer Erwünschtheit in Form einer Überschätzung der Compliance. Weiterhin zeigte sich auch ein Unterschied zwischen den Befragungsmodi, der nicht der postulierten Richtung entsprach. Die Autor*innen stellten hinsichtlich des Tragens einer Maske eine Differenz von rund 14% zwischen den beiden Befragungsmethoden fest. Dabei zeigte sich jedoch, dass Befragte bei der direkten Abfrage zu einem geringeren Anteil die Einhaltung dieser Maßnahme benannten (54,2%) als im List Experiment (67,6%). Diesen gefundenen Effekt interpretieren die Autor*innen als Zeichen, dass Befragte zum Zeitpunkt ihrer Befragung das Tragen einer Maske als *nicht* sozial erwünscht beurteilten (Timmons et al., 2020b). Nichts desto weniger sei ein Effekt der sozialen Erwünschtheit auch bei diesem Item beobachtbar, da sich die Prävalenzen zwischen den Befragten unterschieden. Die Autor*innen schlussfolgern daher auch zum Ende ihres Beitrags, dass eine direkte Abfrage aller untersuchten Items zu verzerrten beziehungsweise künstlichen Complianceraten führen würde und ein List Experiment daher zu empfehlen ist (Timmons et al., 2020a).

Ein weiteres Beispiel für die Verwendung von List Experimenten hinsichtlich coronabezogener Compliance wurde von Larsen et al. (2020) vorgelegt. Die Autoren betrachten die Einhaltung von social distancing, operationalisiert anhand von drei verschiedenen Items, in der dänischen Bevölkerung. Larsen et al. (2020) vergleichen ebenfalls die Ergebnisse zweier List Experimente mit Befragtenantworten auf direkte Fragen. Diese Vergleiche zeigen, im Gegensatz zum erstgenannten Beitrag von Timmons et al., (2020a), keine Unterschiede zwischen den beiden Befragungsmethoden. Die Autoren schlussfolgern daher, dass kein social desirability bias hinsichtlich der genannten Verhaltensweisen vorliegt und daher auch direkte Fragetechniken zu verlässlichen Erkenntnissen zum Ausmaß der Compliance führen.

Als letztes Beispiel für die Verwendung von List Experimenten sei die ländervergleichende Untersuchung von Becher et al. (2020) genannt. Die Autoren erhoben in neun ausgewählten Ländern (Australien, Großbritannien, Deutschland, Frankreich, Italien, Neuseeland, Österreich, Schweden, Vereinigte Staaten von Amerika) Daten mittels einer Onlinebefragung. Die zentrale Frage zum social distancing lautete dabei: *„Ich habe mich mit zwei oder mehr Freunden oder Bekannten getroffen, die nicht in meinem Haushalt leben“*. Die Autoren verglichen die jeweiligen Kontroll- sowie Experimentalgruppen der List Experimente pro Land und bildeten auf diese Weise Angaben für die länderspezifische Compliance bezüglich der zum Zeitpunkt der Befragung geltenden Regelungen. In Deutschland, Schweden sowie Österreich resultierte der Vergleich in einem Anteil von 42% bis 64% von Befragten, die gegen die Vorschrift, im Falle Schwedens gegen die Empfehlung, der Kontaktbeschränkung handelten. In Australien (33,6%), den Vereinigten Staaten (20,9%), Frankreich (12,5%) und Neuseeland (12%) gaben die Befragten in der Minderheit an, Freunde oder Verwandte getroffen zu haben, wenngleich sich auch das Ausmaß dieses Minderheitsverhaltens zwischen den Ländern deutlich unterschied. In Italien und Großbritannien waren keine signifikanten Prävalenzen für einen Verstoß gegen Kontaktbeschränkungen nachweisbar. Die Autoren nehmen diese Werte als Ausgangspunkt für weitere Analysen hinsichtlich des Zusammenhanges der Compliance mit anderen, möglichen Einflussfaktoren. Dabei gehen Becher et al. (2020) davon aus, dass die gewählte Vorgehensweise mögliche Verzerrungen aufgrund sozial erwünschten Antwortverhaltens sowie Messfehler verringern konnte.

Wie diese exemplarische Auflistung hinsichtlich der Verwendung von List Experimenten zeigt, wurde diese Technik von verschiedenen Autor*innen als nützliches Tool für die Erhebung coronabezogener Compliance betrachtet und führte zum Teil auch zu einer veränderten Schätzung hinsichtlich des Ausmaßes der verschiedenen Verhaltensweisen. Im Falle der dänischen Studie führte die Verwendung des List Experiments zwar nicht zu veränderten Prävalenzen, gleichzeitig wurden aber auch keinerlei negative Auswirkungen dieser Befragungstechnik festgestellt. In den beiden anderen Studien konnten erhöhte Verstöße gegen verschiedene präventive Maßnahmen zur Eindämmung der Pandemie festgestellt werden. Unter der Annahme, dass coronabezogene Compliance als sensibles Thema zu verzerrten sowie eher sozial erwünschten Antworten von Befragten führt, kann diese Veränderung als wahrscheinlich eher realistisches Abbild der Realität betrachtet werden.

Zusammenfassend kann somit die Verwendung von List Experimenten als sinnvolle Ergänzung betrachtet werden, die potenziell zu einer Verbesserung der Schätzung von Compliance führt.

Parallelen bei der Erhebung von Korruptionsdelikten sowie der Einhaltung von Corona-Maßnahmen

Bei der Erhebung coronabezogener sowie korruptionsbezogener Compliance können Gemeinsamkeiten benannt werden, welche in der Folge auch eine Verwendung von List Experimenten für die Erhebung korrupter Verhaltensweise nahelegen. Diese Gemeinsamkeiten sollen im Folgenden dargestellt werden.

Der Fokus auf (vergangene oder generelle) Verhaltensweisen von Befragten – genauer: einerseits die Einhaltung von Anti-Corona-Maßnahmen sowie andererseits die Befolgung von Antikorruptionsnormen oder Integritätsregeln – stellt eine offensichtliche Gemeinsamkeit zwischen beiden Themenkomplexen in Befragungen dar.

Weiterhin verweisen die erfragten Verhaltensweisen beider Themenkomplexe auf gesellschaftliche Normen. Diese Normen sind dabei zum einen langfristig gewachsen und in der Gesellschaft stark verankert, wie im Fall des gesellschaftlichen Ideals keine korrupten Handlungen zu begehen. Zum anderen bestehen die Verhaltensnormen im Zusammenhang mit COVID-19 (Hygienemaßnahmen, Abstandsgebot) erst für einen – relativ – kurzen Zeitraum. Dennoch führen die gesetzgeberischen Vorschriften, die nur teilweise kontrolliert oder durchgesetzt werden (können), zu starken gesellschaftlichen Erwartungen an das individuelle Verhalten jedes Einzelnen. Dies führt in der Folge zu einer starken normativen Prägung beider Befragungsthemen. Sowohl bei der Beantwortung von entsprechenden Items zum Themenbereich der Korruption als auch zur Einhaltung verschiedener Maßnahmen zur Eindämmung der COVID-19 Pandemie spielt die soziale Erwünschtheit eine zentrale Rolle. In beiden Fällen profitiert gewissermaßen eine kleine(re) Gruppe gegenüber der Gesellschaft, wenn etwa eine korrupte Praktik angewandt oder gegen eine Auflage zur Eindämmung der Pandemie verstoßen wird. Diese Vorteilsnahme (im weitesten Sinne) muss geheim gehalten werden. Bezüglich der Einhaltung von Anti-Corona-Maßnahmen sind informelle sowie formelle Sanktionen denkbar, wenn Befragte einen Verstoß in einem Survey zugeben. Im Falle

von Korruption wird dieser Umstand sogar noch verstärkt aufgrund möglicher strafrechtlicher Sanktionen beim „Gestehen“ eines korrupten Verhaltens oder von Bestechlichkeit.

Forschende sichern im Rahmen von Umfragen den Befragten die Anonymität der Antworten häufig recht explizit zu und versichern, dass die Antworten nicht an Dritte weitergegeben werden. Auf diese Weise soll die Angst vor möglichen Sanktionen reduziert sowie die Aussicht auf verlässliche Antworten erhöht werden. Die Surveyforschung konnte bereits in verschiedenen Fällen sowie bezüglich verschiedener Themenkomplexe nachweisen, dass diese Zusicherungen nicht immer ausreichend sind, damit Befragte auch tatsächlich sozial unerwünschte Meinungen oder sozial unerwünschtes Verhalten zugeben. Das sozial nicht erwünschte Verhalten müsste hinsichtlich beider hier thematisierter Fragenkomplexe verheimlicht werden. Dies führt in der Folge dazu, dass ein/e Befragte/r im Falle eines Verstoßes gegen Anti-Corona-Maßnahmen oder Anti-Korruptionsmaßnahmen ebendieses Verhalten nicht zugeben möchte.

Laut Tourangeau und Yan (2007) kann die Angst vor Aufdeckung eigener Verhaltensweisen (oder Meinungen) von Befragten als heikler oder sensibler Inhalt von Fragen kategorisiert werden. In diesem Sinne können sowohl Fragen nach korrupten Verhaltensweisen von Befragten als auch coronabezogene Compliance als sensibel eingestuft werden. Die Sensitivität der entsprechenden Frage ist dabei potenziell abhängig von der Antwort der/des Befragten (Tourangeau & Yan, 2007). Im Falle von Korruption ist die Frage zu vergangenen korrupten Verhaltensweisen nur für den Anteil der Befragten sensibel, welche derartige Taten begangen haben. Für den Anteil der Befragten, die dies nicht getan haben, dürfte ein entsprechendes Item in einem Fragebogen hingegen relativ unproblematisch sein, solange sie die Frage nicht als gesellschaftliches Tabu wahrnehmen. Im Falle eines solchen Tabus würde allgemein nicht offen über das betreffende Thema gesprochen, wodurch auch eine Thematisierung während einer Befragung als unangemessen und übergriffig wahrgenommen werden kann. Im Falle eines solchen Themengebietes würden Befragte ebenfalls potenziell die Beantwortung einer Frage verweigern oder mit editierten Antworten der Frage ausweichen. Wenngleich die beiden genannten Kategorien eher fließend ineinander übergehen sowie sich bezüglich eines konkreten Fragebogenitems die beiden Aspekte potenziell vermischen, kann für die hier thematisierten Themengebiete davon ausgegangen werden, dass sich die Angst vor Entdeckung stärker auf die Antworten von Befragten auswirkt.

Daraus lassen sich zweierlei vorläufige Schlussfolgerungen ableiten, die sowohl für Korruptionsthemen als auch für Themen mit COVID-19-Bezug gelten: Einerseits ist von einem sogenannten social desirability bias hinsichtlich beider Themenkomplexe auszugehen. In Studien zur Einhaltung von Anti-Corona-Maßnahmen berichteten Befragte hohe Regelkonformität (Barari et al., 2020; Perrotta et al., 2020). Aufgrund der sozialen Erwünschtheit, sich mit den Regelungen übereinstimmend zu verhalten, könnte ein Anteil dieser Selbstauskünfte möglicherweise einer sozial erwünschten Antwort anstelle einer wahrheitsgemäßen Auskunft zum eigenen Verhalten entsprechen. Gleiches gilt für korruptes Verhalten und diese Antwortstrategie wird aufgrund der befürchtenden strafrechtlicher Sanktionen und der Geheimhaltungserwartung von Mittätern verstärkt. Aus dem Gesagten lässt sich als weitere Gemeinsamkeit der beiden Befragungsthemen ableiten, dass jeweils von einem substantiellen Dunkelfeld auszugehen ist. Die Existenz dieses Dunkelfelds bedeutet, dass erhobene Daten potenziell ein verzerrtes Abbild der Realität darstellen können. Im Falle von Korruption und dem Verstoß gegen Maßnahmen zur Pandemiebekämpfung ist dabei von einer Unterschätzung des Ausmaßes dieser Verhaltensweisen aus den bereits genannten Gründen auszugehen.

Ausblick

Zusammenfassend wurden im Laufe dieses Beitrages somit sowohl die Einhaltung von präventiven Maßnahmen der Eindämmung der Corona-Pandemie als auch Compliance im Sinne der Vermeidung von Korruption als sensitive Themen diskutiert. Die Sensitivität der beiden Befragungsthemen führt in der Folge zu potenziell verzerrten Daten im Falle einer klassischen Abfrage in einem Interview oder einer Umfrage. Dies liegt in dem Umstand begründet, dass Befragte diese Themen nicht oder nicht wahrheitsgemäß beantworten wollen, da sie diese (1) als gesellschaftliches Tabu verstehen, (2) ihre wahre Antwort aufgrund einer Angst vor Aufdeckung nicht preisgeben möchten oder (3) eher sozial erwünscht antworten als wahrheitsgemäß. Diese drei Aspekte können zu verzerrten Daten bei einer einfachen Abfrage führen. Gleichwohl können Forschende bei vorliegenden Daten nicht per se erkennen, ob es sich um (k)eine wahrheitsgemäße Antwort von individuellen Befragten handelt. Potenziell sensitive Themen bergen somit immer das Risiko verzerrter Daten.

In diesem Beitrag wird zur Minimierung dieses Risikos vorgeschlagen, List Experimente bei der Erhebung zu verwenden. Im Rahmen dieser List Experimente kann die Anonymität für die Befragten erhöht werden, wodurch in einigen Studien zur coronabezogenen Compliance das Antwortverhalten der Befragten nachweislich verändert wurde.

Aufgrund der dargestellten Gemeinsamkeit von coronabezogener sowie korruptionsbezogener Compliance stellen List Experimente ebenfalls eine lohnenswerte Erhebungstechnik für Korruptionssitems dar. Mit Hilfe von List Experimenten kann die Wahrscheinlichkeit für wahrheitsgemäße Antworten von Befragten, aufgrund der gesteigerten Anonymität, potenziell erhöht werden. List Experimente können also ein Instrument für die Verringerung des Dunkelfeldes im Bereich der Korruptionsforschung sein und sollten daher für zukünftige Forschungsprojekte in Betracht gezogen werden.

Literatur

- Andersen, H., & Mayerl, J. (2017). Social desirability and undesirability effects on survey response latencies. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 135(1), 68–89.
- Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the covid- 19 epidemic? *The Lancet*, 395(10228), 931–934.
- Barari, S., Caria, S., Davola, A., Falco, P., Fetzer, T., Fiorin, S., Hensel, L., Ivchenko, A., Jachimowicz, J., King, G., Kraft-Todd, G., Ledda, A., MacLennan, M., Mutoi, L., Pagani, C., Reutskaja, E., & Slepoi, F. R. (2020). Evaluating covid-19 public health messaging in Italy: Selfreported compliance and growing mental health concerns. *medRxiv* doi: 10.1101/2020.03.27.20042820.
- Becher, M., Stegmueller, D., Brouard, S., & Kerrouche, E. (2020). Comparative experimental evidence on compliance with social distancing during the COVID-19 pandemic. *medRxiv* doi: 10.1101/2020.07.29.20164806.

- Beck, L., & Nagel, V. (2019). Korruption aus ökonomischer Perspektive. In P. Graeff & T. Rabl (Hrsg.), *Was ist Korruption? Begriffe, Grundlagen und Perspektiven gesellschaftswissenschaftlicher Korruptionsforschung* (S. 39–58). Nomos.
- Betsch, C. (2020). How behavioural science data helps mitigate the COVID19 crisis. *Nature Human Behaviour*, 4, 438.
- Blair, G., & Imai, K. (2012). Statistical analysis of list experiments. *Political Analysis*, 20(1), 47–77.
- Böckenholt, U. (2014). Modelling motivated misreports to sensitive survey questions. *Psychometrika*, 79(3), 515–537.
- Brouard, S., Vasilopoulos, P., & Becher, M. (2020). Sociodemographic and psychological correlates of compliance with the covid-19 public health measures in France. *Canadian Journal of Political Science*, 53(2), 253–258.
- Coutts, E., & Jann, B. (2011). Sensitive questions in online surveys: Experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT). *Sociological Methods and Research*, 40(1), 169–193.
- Daoust, J.-F., Nadeau, R., Dassonneville, R., Lachapelle, E., Bélanger, E., Savoie, J., & van der Linden, C. (2020). How to survey citizens' compliance with COVID-19 public health measures? evidence from three survey experiments. *Journal of Experimental Political Science*. doi:10.1017/XPS.2020.25.
- Glynn, A. N. (2013). What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opinion Quarterly* 77, 159–172.
- Graeber, D., Schmidt-Petri, C., & Schröder, C. (2020a). Hohe Impfbereitschaft gegen Covid-19 in Deutschland, Impfpflicht bleibt kontrovers, *SOEP papers on Multidisciplinary Panel Data Research*, No. 1103. Deutsches Institut für Wirtschaftsforschung (DIW).
- Graeber, D., Schmidt-Petri, C., & Schröder, C. (2020b). Covid-19: Mehrheit der Deutschen würde sich freiwillig impfen lassen, die Hälfte ist für eine Impfpflicht. *DIW aktuell*, No. 54. Deutsches Institut für Wirtschaftsforschung (DIW).

- Graeff, P., & Dombois, R. (2019). Soziologische Zugänge zur Korruptionsproblematik. In P. Graeff & T. Rabl (Hrsg.), *Was ist Korruption? Begriffe, Grundlagen und Perspektiven gesellschaftswissenschaftlicher Korruptionsforschung* (S. 123–142). Nomos.
- Graeff, P., & Steßl, A. (2017). Effektive Compliance: Ursachen, Hindernisse und Lösungsvorschläge. In C. Stark (Hrsg.), *Korruptionsprävention: Klassische und ganzheitliche Ansätze*. Springer Fachmedien.
- Hipp, L., Bünning, M., Munnes, S., & Sauermann, A. (2020). Problems and pitfalls of retrospective survey questions in COVID-19 studies. *Survey Research Methods*, 14(2), 109–114.
- Huebner, M., Spieß, C. K., Siegel, N. A., & Wagner, G. G. (2020). Wohlbefinden von Familien in Zeiten von Corona: Eltern mit jungen Kindern am stärksten beeinträchtigt. *DIW Wochenbericht*, No. 30+31. Deutsches Institut für Wirtschaftsforschung (DIW).
- Imai, K. (2011). Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association*, 106(494), 407–416.
- Jann, B., Krumpal, I., & Wolter, F. (2019). Editorial: Social desirability bias in surveys collecting and analyzing sensitive data. *methods, data, analyses*, 13(1), 3–6.
- Juhl, S., Lehrer, R., Blom, A. G., Wenz, A., Rettig, T., Reifenscheid, M., Naumann, E., Möhring, K., Krieger, U., Friedel, S., Fikel, M., & Cornesse, C. (2020). Die Mannheimer Corona-Studie: Gesellschaftliche Akzeptanz politischer Maßnahmen und befürchtete Konsequenzen für die Wirtschaft. https://madoc.bib.uni-mannheim.de/55140/1/2020-0403_Wochenbericht.pdf. Zugegriffen: 15. Apr. 2021.
- Kapteyn, A., Angrisani, M., Bennett, D., Bruine de Bruin, W., Darling, J., Gutsche, T., Liu, Y., Meijer, E., Perez-Arce, F., Schaner, S., Thomas, K., & Weerman, B. (2020). Tracking the effect of the COVID-19 pandemic on the lives of American households. *Survey Research Methods*, 14(2), 179–186.
- Kühne, S., Kroh, M., Liebig, S., & Zinn, S. (2020). The need for household panel surveys in times of crisis: The case of SOEP-CoV. *Survey Research Methods*, 14(2), 195–203.
- Kuha, J., & Jackson, J. (2014). The item count method for sensitive survey questions: Modelling criminal behaviour. *Journal of the Royal Statistical Society*, 63(2), 321–341.

- Kreuter, F., Barkay, N., Bilinski, A., Bradford, A., Chiu, S., Eliat, R., Fan, J., Galili, T., Haimovich, D., Kim, B., LaRocca, S., Li, Y., Morris, K., Presser, S., Sarig, T., Salomon, J. A., Stewart, K., Stuart, E. A., & Tibshirani, R. (2020). Partnering with Facebook on a university-based rapid turn-around global survey. *Survey Research Methods*, 14(2), 159–163.
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys. A Literature Review. *Quality & Quantity*, 47(4), 2025–2047.
- Krumpal, I., & Voss, T. (2020). Sensitive questions and trust: Explaining respondents' behavior in randomized response surveys. *SAGE Open*. doi:10.1177/2158244020936223.
- Larsen, M., Nyrup, J., & Petersen, M. B. (2020). Do survey estimates of the public's compliance with COVID-19 Regulations suffer from social desirability bias? *Journal of Behavioral Public Administration*, 3(2), 1–9.
- Morin-Chassé, A., Bol, D., Stephenson, L. B., & St-Vincent, S. L. (2017). How to survey about electoral turnout? The efficacy of the face-saving response items in 19 different contexts. *Political Science Research and Methods*, 5(3), 575–584.
- Munzert, S., & Selb, P. (2020). Can we directly survey adherence to non-pharmaceutical interventions? Evidence from a list experiment conducted in Germany during the early corona pandemic. *Survey Research Methods*, 14(2), 205–209.
- Näher, A.-F., & Krumpal, I. (2012). Asking sensitive questions: The impact of forgiving wording and question context on social desirability bias. *Quality & Quantity*, 46(5), 1601–1616.
- Naumann, E., Mata, J., Reifenscheid, M., Möhring, K., Wenz, A., Rettig, T., Lehrer, R., Krieger, U., Juhl, S., Friedel, S., Fikel, M., Cornesse, C., & Blom, A. G. (2020). Die Mannheimer Corona-Studie: Schwerpunktbericht zum Angstempfinden in der Bevölkerung. http://www.uni-mannheim.de/media/Einrichtungen/gip/Corona_Studie/Schwerpunktbericht_Angstempfinden_Mannheimer_Corona_Studie.pdf. Zugegriffen: 7. Apr. 2021.
- Neumann-Böhme, S., Varghese, N. E., Sabat, I., Barros, P. P., Brouwer, W., van Exel, J., Schreyögg, J., & Stargardt, T. (2020). Once we have it, will we use it? A European survey on willingness to be vaccinated against COVID-19. *The European journal of health economics*, 21(7), 977–982.

- Perrotta, D., Grow, A., Rampazzo, F., Cimentada, J., Del Fava, E., GilClavel, S. & Zagheni, E. (2020). Behaviors and attitudes in response to the covid-19 pandemic: Insights from a cross-national facebook survey. *medRxiv*. doi: 10.1101/2020.05.09.20096388.
- Post, J. C., Class, F., & Kohler, U. (2020). Unit nonresponse biases in estimates of SARS-CoV-2 prevalence. *Survey Research Methods*, 14(2), 115–121.
- Richards, L. (2017). Using Survey Methods to Research Corruption. In A. K. Schwickerath, A. Varraich & L.-E. Smith (Hrsg.), *How to research corruption? Conference Proceedings: Interdisciplinary Corruption Research Forum*, June 2016, 4–12.
- Rosenfeld, B., Imai, K., & Shapiro, J. (2016). An empirical validation study of popular survey methodologies for sensitive questions. *American Journal of Political Science*, 60(3), 783–802.
- Sibley, C. G., Greaves, L. M., Satherley, N., Wilson, M. S., Overall, N. C., Lee, C. H. J., Milojeć, P., Bulbulia, J., Osborne, D., Milfont, T. L., Houkamau, C. A., Duck, I. M., Vickers-Jones, R., & Barlow, F. K. (2020). Effects of the COVID-19 pandemic and nationwide lockdown on trust, attitudes toward government, and well-being. *American Psychologist*, 75(5), 618–630.
- Skarbek-Kozietulska, A., Preisendörfer, P., & Wolter, F. (2012). Leugnen oder gestehen? Bestimmungsfaktoren wahrer Antworten in Befragungen. *Zeitschrift für Soziologie*, 41(1), 5–23.
- Timmons, S., McGinnity, F., Belton, C., Barjaková, M., & Lunn, P. (2020a). It depends on how you ask: Measuring bias in population surveys of compliance with COVID-19 public health guidance. *Journal of Epidemiology and Community Health*, 75, 387–389.
- Timmons, S., McGinnity, F., Belton, C., Barjaková, M., & Lunn, P. (2020b). It depends on how you ask: Measuring bias in population surveys of compliance with COVID-19 public health guidance. *Supplementary Material*. doi:10.1136/jech-2020-215256.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859–883.

- Trappmann, M., Krumpal, I., Kirchner, A., & Jann, B. (2014). Item sum. A new technique for asking quantitative sensitive questions. *Journal of Survey Statistics and Methodology*, 2(1), 58–77.
- Utych, S. M., & Fowler, L. (2020). Age-based messaging strategies for communication about COVID-19. *Journal of Behavioral Public Administration*, 3(1), 1–14.
- Wolter, F. (2019). A New Version of the Item Count Technique for Asking Sensitive Questions Testing the Performance of the Person Count Technique. *methods, data, analyses*, 13(1), 169–192.
- Wolter, F., & Laier, B. (2014). The effectiveness of the item count technique in eliciting valid answers to sensitive questions. An evaluation in the context of self-reported delinquency. *Survey Research Methods*, 8(3), 153–168.
- Wolter, F., & Preisendörfer, P. (2013). Asking sensitive questions: An evaluation of the randomized response technique versus direct questioning using individual validation data. *Sociological Methods & Research*, 42(3), 321–353.

3 SECOND PAPER

3.1 English Transcript

Vignette studies

Peter Graeff & Franziska Dunkelmann

Introduction

Collecting data on corruption in surveys or interviews is difficult for a variety of reasons. As a reader of this article, please imagine that you are called by a survey institute and asked whether you have committed corruption in the last five years.

Usually, interviewees would first want to categorise what is meant by the term "corruption". There is no consistent definition and different offences fall under this term in criminal law (Graeff, 2019; Niehaus, 2019). Some interviewees may believe that tax evasion is a form of corruption. Others may consider not registering their cleaner to be a corrupt offence. Without a clear specification of what is meant by the term "corruption", some respondents would answer such a question with idiosyncratic ideas based on their understanding of corruption which may differ from the understanding of others.

It is possible that this brief 'thought experiment' may raise further questions for you, the reader: Are the interviewers implying that you are committing or could commit criminal offences? If you had actually committed corruption offences, why should you answer honestly? Corruption is a criminal offence. Why should you incriminate yourself?

How meaningful can data from such a survey be, if many respondents – potentially the actual perpetrators – do not answer or answer incorrectly while the others are accused of being potential perpetrators?

Despite these difficulties, one survey paradigm is becoming increasingly popular in empirical corruption research: the vignette-based factorial survey¹⁷ (Auspurg & Hinz, 2015). It combines a survey with the advantages of an experiment in that the respondents are given short descriptions of situations (known as vignettes), whose key statements are (experimentally) modified and randomly assigned to the respondents.

This article begins with an overview of the distinctive characteristics of the data collection on corrupt offences and an examination of the strategies employed by previous research to collect data on corruption at the individual level. Subsequently, the characteristics and special features of a vignette-based factorial survey are outlined. The questions raised in the "thought experiment" will be linked to the answers from previous research, which have primarily emerged as best practice approaches. Subsequently, some studies are presented in which a survey of certain corruption offences is scientifically usable. The concluding chapter highlights some of the advantages and weaknesses of the vignette-based factorial survey method, as well as its potential for research.

Special features of data collection on corruption offences

Corruption is considered a sensitive topic when raised in surveys (Dunkelmann, 2022, p. 171; Agerberg, 2022, p. 674). One obvious reason may be that the topic has negative connotations in Westernised countries and is seen as something reprehensible. Even if respondents do not know exactly what defines the crime of corruption, the information that corruption is a criminal offence is sufficient for respondents to reconsider their answers and possibly give answers that do not correspond to their own actions or attitudes. With regard to the sensitivity of corruption, Delios et al. (2024, p. 238) identify a systematic survey error, especially in surveys with information about one's own behaviour or attitude: *"Legal sanctions against corruption and cultural norms that discourage bribery will impel respondents to underreport or outright lie on the survey, leading corruption reports to be biased downwards."*

According to Tourangeau and Yan (2007, p. 860), three dimensions of sensitivity can be distinguished in the context of surveys. The first dimension relates to topics with which

¹⁷ In this article, we mainly focus on vignette-based factorial survey studies. In addition to these, there are a number of similar types of studies, such as simple vignette studies, which are frequently employed by psychologists.

respondents associate social taboos. This can certainly apply to questions about corruption. If these topics are addressed in surveys, these questions can be perceived as intrusive (as may have been the case for you in the "thought experiment" at the beginning of the article). Such taboo topics are not discussed in everyday life, and respondents therefore perceive questions about them as inappropriate. According to Tourangeau and Yan (2007, p. 860), respondents will see such questions *"as an invasion of privacy, regardless of what the correct answer for the respondent is"*.

The second dimension of sensitivity relates to questions where the disclosure of certain answers poses a threat. Such threats relate to answers that could have negative consequences, for instance if the respondent's information becomes known to a third party, such as other persons or authorities. Questions about the (previous) corrupt behaviour of respondents are for instance dangerous from the perspective of perpetrators. Perpetrators have no interest in their offences becoming public. Since interviewees without corrupt behaviour have nothing to hide, there is no risk of disclosure. This example illustrates that this dimension of sensitivity can lead to biased data on corrupt behaviour in a direct survey, as respondents who have committed acts of corruption in the past have different incentives to answer honestly than respondents who have not committed acts of corruption in the past.

The last of the three dimensions of sensitivity according to Tourangeau and Yan (2007) relates to socially desirable response behaviour. This means that respondents tend to answer in accordance with the social norms of their society or certain groups (Tourangeau et al., 2000). According to Tourangeau and Yan (2007, p. 860), certain conditions must be met: *"This conception of sensitivity presupposes that there are clear social norms regarding a given behavior or attitude; answers reporting behaviors or attitudes that conform to the norms are deemed socially desirable, and those that report deviations from the norms are considered socially undesirable."* This dimension can also lead to biased data if respondents behaved undesirable or their attitude violates the social norm. Norm conforming respondents have no incentive to edit their answer. In Western countries, there is also a high normative pressure not to present oneself as corrupt due to the criminal nature of corrupt behaviour. This may be particularly pertinent given that instances of corruption are most prevalent in professional settings. Consequently, individuals who have engaged in corrupt practices would also make disparaging remarks about their work. The last considerations suggest that social desirability

can be regarded as a response strategy on the one hand. On the other hand, social desirability has been considered and analysed as an individual response tendency (in the sense of a stable character trait) since early studies on this topic (e.g., DeMaio, 1984). It is unlikely that corruption is perceived as a neutral or even socially desirable behaviour in Western countries, regardless of how social desirability is characterised. This is also due to the fact that in corrupt situations, a relatively preconditional illegal exchange takes place. In addition to its social and economic damage, is also considered morally wrong (Persson et al., 2013, pp. 455f).

In German criminal law, this exchange is associated with the fact that corruption presupposes voluntariness to commit the offence and consensus among the perpetrators (Niehaus, 2019). This presuppositional exchange is relevant for the social sciences because it requires a special relationship between the perpetrators that does not occur in this form in other criminal offences. If a corrupt actor were involuntarily pressurised into an exchange of corruption, this would constitute blackmail or coercion, which suggests other social interests and incentives. Anyone who is blackmailed or coerced can contact the police or law enforcement authorities, file a complaint, and initiate a criminal prosecution process. Typically, corrupt partners do not act like this: they do not report each other after a successful act of corruption, as this would make their offence public and invalidate their own corrupt advantage. On a social level, corruption presupposes that the parties agree on what their act of corruption consists of and that this has to be kept a secret. There is a "wrongful agreement" (Niehaus, 2019, p. 228) that cannot be forced.

A counter clerk may well refuse an offer of a bribe in the form of an amount of money and refrain from issuing a driving licence. He would refrain from accepting a bribe if he were aware that the bribe-giver would subsequently disclose the act of corruption to the public. Conversely, a bribe-giver would not offer money if the corruption was later revealed and the driving licence was then withdrawn.

There are several paragraphs in German criminal law that are associated with corruption offences. These are generally not known to interviewees. There are also grey areas where the classification is not clear, especially for amateurs. What is labelled as corruption in direct survey items or predefined vignettes therefore has an influence on the respondents' thought and response processes. In empirical studies the bribery paradigm dominates as a representation of corruption in direct surveys as well as in experimental games. In this way,

no particular subtypes or grey area examples of corrupt acts are investigated. Instead, the bribery offences clearly reflect corruption content that is also understood as corruption by the general public. Interestingly, this applies not only to Germany, but also to most other countries, even if a high level of corruption predominates there (Fisman & Golden, 2017, pp. 26ff).

The literature on the empirical implementations on the topic of corruption contains several facets of what exactly should be investigated and possibly also explained. These primarily include whether respondents have an intention to commit corrupt acts in the future (*intention to act*), whether they have already committed corrupt offences in the past (for instance in the last three years; *prevalence*), what their assessment and attitude towards acts of corruption is (*attitude*) or how they assess (the frequency and type of) acts of corruption committed by others, for instance in their company or authority (*external observation*).

Oliveros and Gingerich (2020, p. 384) distinguish two reasons why deliberately misreporting can occur in corruption surveys in particular: on the one hand, there are factors related to the survey itself (such as a lack of anonymity in the survey), and on the other hand, there are factors related to the respondent themselves (such as threat of disclosure).

In recent decades, some survey research has focussed on the question of how the anonymity of respondents can be ensured so that they can give a "true" answer without having to deal with negative consequences later on. A number of techniques have been developed to ensure anonymity. These include, for instance, the randomised response technique (e.g., Krumpal & Voss, 2020) or non-randomised response techniques such as the crosswise model (e.g., Oliveros & Gingerich, 2020, p. 388). However, the application of these techniques requires a number of assumptions (e.g., that the respondents trust the researcher to apply the techniques correctly), so that in practice the desired results are not always achieved if these assumptions do not apply (Lensvelt-Mulders & Boeijs, 2007; Boeijs & Lensvelt-Mulders, 2002, p. 31).

When investigating corruption with a vignette-based factorial survey, an attempt is made to minimise sensitivity via the presentation style and the indirect question type (Sauer et al., 2020, p. 197). How this is done is described in more detail in the next chapter.

The vignette-based factorial survey

A vignette-based factorial survey combines the real-life and contextual nature of vignettes with an experimental design (Auspurg & Hinz, 2015). Vignettes are short, fictional descriptions of situations or people. As part of the experimental design, different characteristics of these variables (known as levels) are randomly varied in the survey. Different respondents are thus presented with vignettes that each contain different characteristics of the variables. The respondents are then asked to give answers to these fictional descriptions.

Vignette-based factorial surveys are considered useful in research situations in which observations of the social phenomenon or experiments under real conditions would not be possible or would be unethical (Auspurg & Hinz, 2015). Corrupt situations fall into this category. At present, corrupt deals between perpetrators cannot be observed, nor can bribery with consequences be carried out experimentally (e.g., with appropriate criminal prosecution). In experimental games in which participants recreate corrupt situations and deal with real gains or losses (e.g., in fictitious money) (for instance Abbink, 2002 or Banerjee & Mitra, 2018), the consequences for corrupt game players end the end of the game. Nevertheless, experimental games can explore the mechanisms that play a role in decisions about corruption.

Vignette-based factorial surveys can explore mechanisms of corruption in the same way due to the experimental variation of the explanatory variables (Graeff, 2016). In terms of researching the occurrence of or the fight against corruption, the vignettes contain the variables that can be expected to influence the occurrence of or the fight against corruption on the basis of theoretical or empirical guidelines. The amount of the monetary benefit for a bribe-taker or the probability of being discovered in the course of corruption can be examples. The scenarios put the respondents' answers into an understandable context if the vignettes are realistic ("real-life stories", Sauer et al., 2020, p. 197). Several explanations or causes of corruption can be analysed at the same time, thus taking into account the multi-dimensional phenomenon of corruption. It is also advantageous that the respondents are given a similar basic scenario, meaning that the general context of their answers remains comparable.

In terms of data collection, the combination of experimental design and survey means that large numbers of cases (especially at the vignette level) can be realised (Aguinis & Bradley,

2014), thus increasing test power. In addition, generalisations of the results can be derived, which is usually rarely possible with purely experimental research studies in the social sciences (Mutz, 2011, p. 22). The experimental design has additional advantages that do not exist with simple surveys. These include the fact that random assignments of the explanatory variable values make these variables independent of each other and their individual influences can therefore also be shown.

As already indicated, vignettes are also seen as a way of reducing various types of sensitivity in surveys (such as social desirability; Auspurg et al., 2014; Wallander, 2009). Armacost et al. (1991) asked the same sensitive questions directly, via vignettes or via the randomised response method in a postal survey of company managers of manufacturing companies. These questions dealt with questionable economic behaviour, such as the use of insider information or the implementation of fake job interviews. Responses were either collected as a self-report or the perception of the behaviour of others (with the exception of the randomised response survey). Direct questioning provoked more socially desirable answers and at the same time fewer socially undesirable responses in the self-report than the other two methods. The generalisability of the results is questioned by the authors themselves due to the low response rate of 20.2% (Armacost et al., 1991, p. 1084).

A selection effect of those who typically use such corrupt procedures may be reflected in the high unit non-response rate. Jensen et al. (2010), for instance, argue that non-response and biased responses in surveys on the perception of corruption among company employees represent a self-protective mechanism that depends on the political environment. Using World Bank data on 44,000 companies in 72 countries, they show that in politically repressive countries, non-response and biased answers can be assumed to be more frequent in surveys on corruption issues.¹⁸

Other reasons why vignette-based factorial surveys might produce better results than direct questioning with regard to sensitive content are the hypothetical descriptions and the usually

¹⁸ In this study, “*potential false responses*” are calculated from the discrepancy between the survey data and the experts' assessment of the extent of corruption in the countries. However, the study does not ask about the extent of corruption, but whether corruption is a problem and to what extent the severity of the problem is categorised as an obstacle (5-point scale). Respondents who have come to terms with corruption as a necessary practice are unlikely to perceive it as a problem, which reduces the comparability with the expert assessment (Jensen et al., 2010, p. 1485).

multiple explanatory variables (Auspurg et al., 2014, p. 141). In the event that a protagonist is described in the vignette and the interviewees are asked to answer on their behalf, they are merely indirectly expressing their own opinion on this hypothetical situation. Given that the levels of several explanatory variables are typically varied across the different vignettes, the ability to focus on a single (potentially particularly sensitive) variable is reduced (Wallander, 2009, p. 506).

The study by Auspurg et al. (2014), in which respondents assessed the fairness of earnings related to the gender of protagonists, also found evidence that vignette-based factorial surveys can reduce the effect of sensitive topics - compared to the situation in which survey participants are asked directly (Auspurg et al., 2014, p. 143f). Their study also found evidence that the effect of sensitivity decreases in the case of a vignette-based factorial survey, particularly for more highly educated people (Auspurg et al., 2014, p. 145).

Studies about corruption

The vignette-based factorial survey is a method for investigating theoretical specifications or hypotheses with individual data. It requires a theory or at least a well-founded hypothesis in order to select and design the explanatory variables (sometimes also referred to as dimensions) and their characteristics (levels) for the vignettes (Auspurg & Hinz, 2015). To date, there is no theory in corruption research that could comprehensively explain the emergence or maintenance of corrupt practices at the individual level and serve as a basis for the design of a vignette-based factorial survey. Theoretical guidelines are derived from a variety of social science disciplines, including political science, psychology, and sociology, as well as neighbouring disciplines such as law and economics.

This also becomes clear in the following studies which were selected as examples to illustrate the use of vignette-based factorial surveys in corruption research.

Political science studies typically use factorial surveys if they aim to research political corruption. As one example, Vera (2020) varies three aspects in the description of a fictitious Peruvian politician running for congress. The description focuses on whether he has already been accused of corruption offences (or not), whether he has achieved more successes for the local community than the majority of mayors (or not), and whether he is running in a

constituency where the level of corruption is high (or low). The vignettes were used as part of a nationwide survey in Peru. The explained variable measures (as a behavioural intention) the probability assessment of the 1308 Peruvian respondents to vote in favour of the described candidate.¹⁹ Respondents were asked to indicate this probability on a 7-point scale. The results suggest a negative influence of corruption allegations on the probability of voting for the candidate. Due to the multifactorial design, the interactions with other dimensions could be analysed. The results suggest that political candidates who have proven themselves to be competent are less likely to be punished at the polls for corrupt offences.

Based on theoretical considerations from political psychology regarding emotions and with a similar design to Vera (2020), Birch et al. (2017, p. 901) investigate whether certain behaviours that fall within the grey area of corruption lead to a negative evaluation among respondents. The survey was conducted in France and, in contrast to Vera (2020), is therefore an example of a study in a Western country. The described behaviours focused on the French context and included double vs. single mandate, a possible illegitimate political advantage (or the deliberate emphasis that this is not the case) and a possible financial advantage linked to self-interest (or the absence of entanglement with personal gain). The dependent variable measures on an 11-point scale whether the behaviour is acceptable from the respondents' point of view. The results show that illegitimate political advantages and self-interest-related financial advantages are rated significantly less acceptable. Furthermore, Birch et al. (2017, p. 905) link these results with classifications of respondents according to emotion-psychological categorisations of fear and anger. These results are found to be inconsistent.

Sociological research that investigates corruption with vignette-based factorial surveys draws on other theoretical starting points, like corruption as a decision between a particular and universal norm (Schweitzer, 2022) or a decision that is motivated by specific incentives (Graeff, 2016). In accordance with this, the vignettes and explained variables are designed to provide responses pertaining to norms and values (Auspurg & Hinz, 2015), or an intention of the

¹⁹ A special feature in the implementation of the factorial survey can be seen in the experimental design by Vera (2020). Each respondent was randomly assigned exactly one vignette (known as "between-subject design"), which Vera (2020, p. 660) explains as follows: "[...] this experimental approach was adopted as an unobtrusive way to measure the socially desirable attitude of rejecting a corrupt candidate. First, it does not expose a respondent to multiple vignettes that would make them aware of the experimental variations. Second, it presents a vignette with a rather complex hypothetical candidate that has multiple qualities, making it difficult for the respondent to notice the key factors manipulated in the vignette. And, finally, it asks for the attitudes of a third person, allowing the respondent to answer honestly without explicitly stating their preference."

respondents to decide in the place of the character described in the vignette. According to Liebe et al. (2020, p. 163), such norm or value judgements are measured in an almost elegant way because the underlying (theoretical) concepts and contents are captured in the vignettes in their significance to the corresponding situational variables and are not (as is often the case in direct questioning) surveyed as individual items.

The study by Kleinewiese and Graeff (2020) is an example of group-related value judgements regarding corrupt behaviour. The study employed a sample of volunteer firefighters to investigate whether the presence of a team spirit (as a group norm) can lead to a conflict between particularistic and universalistic norms if team members engaged in deviant behaviour. As with the police and the military, the work of volunteer firefighters heavily depends on team cohesion, enabling them to collectively overcome challenges for or attacks on the team. However, this team spirit also suggests a code of silence when a team member should actually be reported for deviant behaviour. Conversely, universalist norms and laws require that any observed behaviour that is legally punishable must be reported. From the perspective of previous studies, there is a conflict here between loyalty to the team and fairness norms given by the law. This dilemma is attributed to whistleblowing in corruption research (Dungan et al., 2019). The firefighters were requested to estimate the probability that the protagonist of the vignettes would report the deviant behaviour of a team member of the volunteer fire department to their supervisor. This dependent variable was measured on an 11-point scale. The results indicate that a strong team spirit can trigger the aforementioned conflict if team members engage in deviant behaviour. In the event of an external attack on the team in such circumstances, these attacks reinforce the code of silence. In the vignettes, the assault by drunken football fans was varied as an attack on or disruption of an operation of a volunteer fire brigade. As there was no original theory of corruption that can serve as a starting point for such studies, an attempt was made in this case to categorise the results within a broader framework of deviance theory. Other studies on deviance and crime take a similar approach and draw on theoretical starting points from the *General Theory of Crime* or the *Situational Action Theory* (e.g., Kleinewiese, 2022).

In a study by Graeff et al. (2014), students were asked to decide whether they would accept bribes based on described scenarios. The participants were assigned to the role of a tutor based on the information provided in the vignette scenarios. The tutor could decide whether

to certify that a student had passed an exam even though they had not. In the vignettes, factors like the bribe amount, potential penalties, the probability of being discovered and the trustworthiness of the student (and thus the probability of success of the corrupt act) were varied. Following the presentation of each vignette, respondents were asked to their level of preparedness (in terms of behavioural intention) to modify the student's grade, indicated on a 10-point scale. The results show that with increasing incentives (such as higher bribes or greater trustworthiness), the willingness to engage in corrupt behaviour also increases. This willingness decreases with a greater probability of detection. It was notable that normative decision-making variables demonstrated a greater explanatory value and explained more variance than the previously mentioned incentives or deterrence factors. This indicates that, in terms of practical anti-corruption strategy, the empowerment of normative decision-making elements (like the importance of fairness norms) may be more effective than simple reward-punishment mechanisms (e.g., Graeff, 2016).

Psychological studies that investigate corruption usually use traits as independent variables and behavioural intentions (for corrupt actions) as dependent variable. They design the vignettes in terms of unethical decision-making behaviour or as an ethical dilemma (Rabl, 2019, p. 79). Similar to the study by Graeff et al. (2014), De Waele et al. (2021) consider the willingness of students to pay bribes.²⁰ The authors use a vignette study in the form of a between-subject design in Belgium, Germany and the Netherlands. The study can be characterized as a between-subject design because each student receives only one (of three variations of the) vignette. The students were requested to imagine themselves as if they were in the initial semester of their degree programme once more, and had failed an examination. Furthermore, the students were told that they had achieved 9.4 out of 20 points, whereby 9.5 points would have led to the score being rounded up and thus to passing the exam. Following this introduction, one of three vignettes is then randomly displayed which describes the following levels of bribery:

²⁰ In contrast to other studies, the dependent variable of willingness to bribe was not asked as one direct question, but was formed as a geometric mean of four variables: *“by asking how likely they were to bribe in this context (likelihood), how justified offering the bribe was (justification), how comfortable they would feel in doing so (affect), and whether offering a bribe would be a mistake (mistake), which was a reverse item for control”* (De Waele et al., 2021, p. 5).

- 1) Emotional request to the lecturer to enter the exam as passed (labelled as *white bribery* by the authors).
- 2) Suggesting a favour to the lecturer (repairing their car) in return for passing the exam (labelled as *grey bribery* by the authors).
- 3) Handing over an envelope containing €500 to the lecturer with a request to reconsider the grade (labelled as *black bribery* by the authors).

In all three cases, the vignette also contains the information that it can already be assumed from rumours that this behaviour leads to successfully passing the exam. De Waele et al. (2021, p. 5) use social value orientation (SVO) as one of two central explanatory variables: "*according to Messick and McClintock (1968), refers to a stable preference for certain outcomes for oneself against others' outcomes, thus capturing the extent to which an individual is intrinsically mainly concerned with personal versus group well-being.*" While no significant effect can be found for the second explanatory variable (public service motivation), SVO increases the willingness to commit corruption (paradoxically, from the authors' view).²¹

In economic studies that use vignette-based factorial surveys to explain corruption, the focus is on incentives and costs as well as individual-related variables. One example is the study by Dickel and Graeff (2018), where entrepreneurs were asked in a realistic scenario whether they would bribe a foreign official to improve the air in a large city with an innovative technology. As in previous studies, the behavioural intention constituted the dependent variable. The participants indicated a significant influence of the expectation of profit and costs as well as the ability to successfully carry out the corruption. These influencing variables are reinforced by the individual tendency to consider the advantages and disadvantages of the opportunity for corruption (deliberation). Contrary to the theoretical expectation (based on common theories of crime) that the probability of detection negatively influences committing a criminal offence, the probability of detection did not prove to be statistically significant in this study.

²¹ Given the empirical findings in which corruption is seen as a discrepancy between group loyalty and fairness norms (Kleinewiese & Graeff, 2020; Dungan et al., 2019; Schweitzer, 2022), this result shows that the tendency towards group loyalty favours bribery.

Discussion and outlook

At the beginning of this article , we asked you to engage in a thought experiment. We asked you to imagine being called by a survey organisation and being asked whether you have committed corruption in the last five years.

If you placed yourself in this fictitious situation when you first read this prompt (in the introduction to this chapter), then you have done what is required of interviewees when they are confronted with a vignette scenario.

The survey mode (telephone survey) and the time period for which you were asked to answer the question (5 years) could have been dimensions in a vignette-based factorial survey, which could have had different levels. For instance, another vignette could have asked about the last year or the last three years. Following the vignette, at least one question would be linked to this scenario. If the vignette or scenario describes a person in whose place a decision is to be made, then the respondent would not answer the question directly, but with a distance to themselves.

As outlined in the previous article, such a design can have advantages in terms of capturing different facets of corruption. However, there are also difficulties. These include the design of plausible and meaningful scenarios. This usually requires prior knowledge of the context to be analysed in order to include relevant, explanatory factors (dimensions) as well as plausible levels in the vignettes. This knowledge should be derived from theoretical considerations or previous studies. This is difficult in corruption research due to the lack of theory (Graeff, 2019, p. 293). As previously demonstrated, this deficit requires theoretical considerations from a variety of scientific disciplines. As a result, social science vignette-based factorial surveys often remain discipline-specific and are not taken up by neighbouring disciplines due to conceptual and theoretical incommensurability.²²

Before being carried out, vignette-based factorial surveys require profound pretests in the target population. It must be ensured that the dimensions (factors) and associated levels are understood and included in the response process. To this end, the pretest must provide indications that the variation of the explanatory factors (the experimental aspect of the

²² Incommensurability means that two theories use such fundamentally different assumptions or concepts that a direct comparison or translation of one theory into the terms of the other is difficult or impossible.

survey) changes the result. If this is not the case, the vignette dimensions and levels must be adjusted accordingly as well as retested. Compared to a *simple* survey with a standardised questionnaire (which may be able to draw on tried and tested items from other surveys), this process is sophisticated and time-consuming. In addition, those using this method should bear in mind that conducting a vignette-based factorial survey usually requires a higher level of cognitive effort from respondents compared to standardised surveys.

The aforementioned challenges are offset by the advantages of recording sensitive topics such as corruption (Dunkelmann, 2022, p. 171). In light of the three dimensions of sensitivity as outlined by Tourangeau and Yan (2007) – namely social taboos, threat of disclosure and social desirability – it is possible that response bias may occur at the individual level despite the use of vignette-based factorial surveys. Furthermore, it is also possible that item-nonresponse may occur. In the case of social desirability, for instance, the response bias manifests itself in the fact that no approval of corrupt behaviour or no corresponding behavioural intention is given in response to the question about the vignette. As a result, the respondents' answers might not vary across the characteristics of the dependent variable. In the extreme case, all respondents give exactly the same answer: there is no approval of corrupt behaviour or no corresponding behavioural intention. If many respondents choose not to answer the question about the vignette at all (item nonresponse), the number of responses that can be analysed is reduced. Nevertheless, the study results presented in the preceding section demonstrate that even sensitive questions generally yield sufficient variation in respondents' answers to permit analysis.

This may also be due to the fact that respondents do not have to answer a corruption-related question directly, but rather assess a hypothetical situation. If one considers the first dimension of Tourangeau and Yan (2007; social taboos) under this premise, the use of vignette-based factorial surveys can lead to a softened perception of the taboo. Respondents could comment on a taboo topic, because they only have to indirectly state their attitude in this regard.

A reduction in sensitivity can also be derived in accordance with the second dimension of Tourangeau and Yan (2007) due to the indirect type of survey. As the previously referred studies demonstrate, respondents are typically not asked about past actions but rather about their behavioural intentions or attitudes in response to a given scenario. Moreover,

interviewees frequently evaluate scenarios that are not directly related to them personally. For instance, this may occur because the behaviour of the protagonists in the vignette is assessed. Consequently, there is a lack of reflection on one's own behaviour. This reduces the threat of disclosure, as no specific offence has to be admitted. This lack of self-disclosure also weakens the threat of disclosure, because no concrete offence can be reported to third parties.

With regard to social desirability - the third dimension according to Tourangeau and Yan (2007) - a declined sensitivity to corruption can also be assumed here in contrast to the direct question. This has already been shown for other topics characterised by social desirability, which is why this mechanism can also be assumed with regard to the behavioural intention and attitudes towards corruption (Auspurg et al., 2014, p.146). Nevertheless, the second approach of socially desirable response behaviour as a stable personality trait (DeMaio, 1984) - possibly in the sense of a sugar-coated self-image - cannot be taken into account in the context of a vignette-based factorial survey. If respondents answer in the direction of social norms if certain dimensions are mentioned (e.g., when describing a bribe) instead of answering in accordance with their personal attitude or actual behavioural intention, this problem remains despite this survey method.

Summarising the statements made so far, the poignancy can be taken from the questions posed at the beginning, which dealt with the allegation of criminal offences, the reason for answering honestly or self-disclose in the case of corruption topics.²³ A central argument is the indirect confrontation with scenarios instead of direct questions. Instead of accusing interviewees of possible acts of corruption, they are asked to assess hypothetical scenarios. If this is combined with a question about behavioural intentions and/or attitudes towards corruption, it does not rule out self-disclosure (e.g., admitting an attitude towards corruption), but the answer is less specifically related to respondents' past actions. Since self-disclosure in this case is not associated with the same negative consequences (threat of disclosure) as admitting an act of corruption in a direct interview, vignettes could also make it easier for

²³ In addition to these questions, the problem of the lacking clarity of the concept of corruption was raised. This problem can be addressed in two ways within the framework of a vignette-based factorial survey:

1. Interviewees can be asked to assess whether the scenario described constitutes corruption.
2. The vignette can describe a specific act of corruption that is not in the so-called "grey area" but is clearly recognised as corruption by respondents (e.g., the acceptance of bribes, Graeff et al., 2014).

offenders to answer honestly. Moreover, intentions or attitudes are not illegal which excludes juristic implications or aspects of criminal prosecution. For the research on criminal phenomena, the results obtained with the help of vignette-based factorial surveys could be significantly less negatively influenced by response bias and item-nonresponse should also occur less frequently.

Vignette-based factorial surveys will continue to be used as a survey method in corruption research in the future. Their advantages outweigh the problems and the effort involved in their design. It is particularly important to emphasise that this method is able to depict the multitude of factors that influence corruption and, above all, to isolate their influence. By presenting realistic scenarios, decisions and norm assessments can be modelled in real-life situations. To a limited extent, this also bridges the gap between abstract assessments of ethical standards and actual behaviour or behavioural intentions in real-life situations. Since there are ethical limits to the observation and investigation of real corrupt behaviour, vignette-based factorial surveys offer an ethically justifiable way of investigating how respondents react to corruption decisions without displaying morally questionable behaviour of their own or exposing themselves to direct risk. In addition, the methods can be adapted to different samples and contexts and can be used to uncover corruption mechanisms, even with respondent groups that are difficult to motivate for alternative survey methods like participating in experimental games, like managers (Dickel & Graeff, 2016). Although some of these advantages positively differentiate these methods from conventional surveys, some advantages also highlight fundamental difficulties in dealing with corruption research. This implies that the hypothetical scenarios remain hypothetical, and that respondents who are aware of this can indicate deviations from their behaviour in real situations. In addition, the type and modes of implementation (which include, for instance, the scenario and question presentation, the number of dimensions as well as vignettes or the possibility of participating in the survey, Leitgöb & Wolbring, 2021) still raise a number of questions that can only be answered in future studies.

References

- Agerberg, M. (2022). Corrupted estimates? Response Bias in Citizens Surveys on Corruption. *Political Behaviour*, 44, 653-678.
- Aguinis, H. & Bradley, K.J. (2014). Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies. *Organisational Research Methods*, 17(4), 351-371.
- Armacost, R.L., Hosseini, J.C., Morris, S.A. & Rehbein, K.A. (1991). An Empirical Comparison of Direct Questioning, Scenario, and Randomised Response Methods for Obtaining Sensitive Business Information. *Decision Science*, 22, 1073-1090.
- Auspurg, K., Hinz, T., Liebig, S. and Sauer, C. (2014). The Factorial Survey as a Method for Measuring Sensitive Issues. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, & P. Sturgis (Eds.), *Improving Survey Methods*, 137-149.
- Auspurg, K., & Hinz, T. (2015). Factorial survey experiments, Sage, Thousand Oaks.
- Banerjee, R., & Mitra, A. (2018). On monetary and non-monetary interventions to combat corruption. *Journal of Economic Behavior and Organisation*, 149, 332-355.
- Boeije, H. & Lensvelt-Mulders, G. J. L. M. (2002). Honest by Chance: A Qualitative Interview Study to Clarify Respondents' (Non-)Compliance With Computer-Assisted Randomised Response. *Bulletin de Methodologie Sociologique*, 75, 24-39.
- Birch, S., Allen, N.J., & Sarmiento-Mirwaldt, K. (2017). Anger, Anxiety and Corruption Perceptions: Evidence from France. *Political Studies*, 65(4), 893-911.
- Delios, A., Malesky, E.J., Yu, S. & Riddler, G. (2024). Methodological errors in corruption research: Recommendations for future research. *Journal of International Business Studies*, 55, 235-251.
- Dickel, P. & Graeff, P. (2016). Applying Factorial Surveys for Analysing Complex, Morally Challenging and Sensitive Topics in Entrepreneurship Research - The Case of Entrepreneurial Ethics. In E.S.C. Berger & A. Kuckertz (eds.), *Complexity in entrepreneurship, Innovation and Technology Research - Applications of Emergent*

- and Neglected Methods FGF Studies in Small Business and Entrepreneurship. Springer, Berlin, 1-20.
- Dickel, P. & Graeff, P. (2018). Entrepreneurs' propensity for corruption: A vignette-based factorial survey. *Journal of Business Research*, 89, 77-86.
- Dunkelmann, F. (2022). Korruption und Corona. Mögliche Gemeinsamkeiten zweier Befragungsthemen. In S. Wolf & P. Graeff (Hrsg.), *Corona und Korruption. Gesellschaftswissenschaftliche Analysen*, Springer VS, doi: 10.1007/978-3-658-35664-4_7.
- De Waele, L., Weißmüller, K.S., & Van Witteloostuijn, A. (2021). Bribery and the Role of Public Service Motivation and Social Value Orientation: A Multi-Site Experimental Study in Belgium, Germany and the Netherlands, *Frontiers in Psychology*, 12:655964, doi: 10.3389/fpsyg.2021.655964.
- Dungan, J. A., Young, L. & Waytz, A. (2019). The Power of Moral Concerns in Predicting Whistleblowing Decisions. *Journal of Experimental Social Psychology*, 85:1-12.
- Fisman, R. & Golden, M.A. (2017). Corruption: What everyone needs to know. New York: Oxford University Press.
- Graeff, P. (2016). Social Mechanisms of Corruption: Analytical Sociology and its applicability to corruption research. *Analyse/Kritik* 37(1), 53-71.
- Graeff, P. (2019). Begriffsverwendungen und Korruptionsverständnisse in der Forschung: Ein Vergleich von Definitionen und Ansätzen. In: P. Graeff & T. Rabl (Hrsg.): *Was ist Korruption? Begriffe, Grundlagen und Perspektiven gesellschaftswissenschaftlicher Korruptionsforschung*. 2. Auflage. Nomos, Baden-Baden, 271-298.
- Graeff, P., Sattler, S., Sauer, C. & Mehlkop, G. (2014). Incentives and prohibitions of abusing academic positions. Determining students tendencies to consent to bribing offers: a vignette study. *European Sociological Review*, 30(2), 230-241.
- Jensen, N. M., Li, Q. & Rahman, A. (2010). Understanding corruption and firm responses in cross-national firm-level surveys. *Journal of International Business Studies*, 41, 1481-1504.

- Kleinewiese, J. (2022). Good people commit bad deeds together. A Factorial Survey on the moral antecedents of situational deviance in peer groups. *Deviant Behaviour*, 43(11), 1420-1431.
- Kleinewiese, J. & Graeff, P. (2020). Ethical decisions between conflicting priorities of legality and group loyalty: Scrutinising the "Code of Silence" among volunteer firefighters with a Vignette-based Factorial Survey. *Deviant Behaviour*, 42(10), 1228-1241.
- Krumpal, I. & Voss, T. (2020). Sensitive questions and trust: Explaining respondents' behaviour in randomized response surveys. *Sage Open*, 10(3), doi:10.1177/2158244020936223.
- Leitgöb, H. & Wolbring, T. (2021). Die Methoden der sozialwissenschaftlichen Datenerhebung im digitalen Zeitalter. Entwicklungen, Möglichkeiten, Herausforderungen, In T. Wolbring, H. Leitgöb, & F. Faulbaum (Hrsg.) *Sozialwissenschaftliche Datenerhebung im digitalen Zeitalter. Schriftenreihe der ASI - Arbeitsgemeinschaft Sozialwissenschaftlicher Institute*. Springer VS, Wiesbaden, doi: 10.1007/978-3-658-34396-5_1.
- Lensvelt-Mulders, G. J. L. M. & Boeije, H. R. (2007). Evaluating compliance with a computer assisted randomised response technique: a qualitative study into the origins of lying and cheating. *Computers in Human Behaviour*, 23, 591-608.
- Liebe, U., Moumouni, I.M. Bigler, C. Ingabire, C. & Bieri, S. (2020). Using factorial survey experiments to measure attitudes, social norms, and fairness concerns in developing countries. *Sociological Methods & Research*, 49(1), 161-192.
- Mutz, D. C. (2011). *Population-Based Survey Experiments*, Princeton University Press, Princeton, doi: 10.1515/9781400840489.
- Niehaus, H. (2019). Annäherungen an einen Korruptionsbegriff des (deutschen) Strafrechts. In: P. Graeff & T. Rabl (Hrsg.): *Was ist Korruption? Begriffe, Grundlagen und Perspektiven gesellschaftswissenschaftlicher Korruptionsforschung*. 2. Auflage. Nomos, Baden-Baden, 221-234.
- Oliveros, V. & Gingerich, D.W. (2020). Lying about Corruption in Surveys: evidence from a Joint Response Model. *International Journal of Public Opinion Research*, 32(2), 384-395.

- Persson, A., Rothstein, B. & Teorell, J. (2013). Why anticorruption reforms fail - systematic corruption as a collective action problem. *Governance: An International Journal of Policy, Administration, and Institutions*, 26(3), 449-471.
- Rabl, T. (2019). Korruption aus der Perspektive der Psychologie. In: P. Graeff & T. Rabl (Hrsg.): *Was ist Korruption? Begriffe, Grundlagen und Perspektiven gesellschaftswissenschaftlicher Korruptionsforschung*. 2. Auflage. Nomos, Baden-Baden, 77-102.
- Schweitzer, H. (2022). Korruption. Eine sozialwissenschaftliche Darstellung und Analyse. Hamburg: Disserta Verlag.
- Tourangeau, R. & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133 (5), 859-883.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Vera, S.B. (2020). Accepting or resisting? Citizen responses to corruption across varying levels of competence and corruption prevalence. *Political Studies*, 68(3), 653-670.
- Wallander, L. (2009). 25 years of factorial surveys in sociology: A review. *Social Science Research*, 38, 505-520.

3.2 German Original

Vignettenstudien

Peter Graeff & Franziska Dunkelmann

Einleitung

Die Erfassung von Korruption in Befragungen oder Interviews ist aus einer Vielzahl von Gründen schwierig. Stellen Sie sich bitte als Leser*in dieses Beitrages vor, Sie würden von einem Befragungsinstitut angerufen und Sie bekämen die Frage gestellt, ob Sie in den letzten fünf Jahren eine Korruption begangen haben.

Üblicherweise würden Befragte zunächst einordnen wollen, was mit dem Begriff „Korruption“ gemeint ist. Es gibt keine einheitliche Definition und im Strafgesetz fallen unterschiedliche Tatbestände unter diesen Begriff (Graeff, 2019; Niehaus, 2019). Einige Befragte könnten meinen, dass Steuerhinterziehung eine Form von Korruption ist. Andere könnten meinen, dass die Nichtanmeldung ihrer Putzkraft einen Korruptionsdelikt darstellt. Ohne eine klare Vorgabe, was mit dem Begriff „Korruption“ gemeint ist, würden einige Befragte aufgrund ihres Verständnisses von Korruption, welches sich vom Verständnis anderer unterscheiden könnte, mit eigentümlichen Vorstellungen auf eine solche Frage antworten.

Möglicherweise ruft das kleine „Gedankenexperiment“ bei Ihnen als Leser*in bereits weitere Fragen auf: Unterstellen die Befragenden Ihnen, dass Sie Straftaten begehen oder begehen könnten? Wenn Sie tatsächlich Korruptionstaten begangen hätten, stellt sich die Frage, warum Sie ehrlich antworten sollten? Korruptionstaten gehören zu den strafrechtlich verfolgten Delikten. Warum sollten Sie sich selbst beschuldigen?

Wie aussagekräftig können Daten aus einer solchen Erhebung sein, wenn viele Befragte – möglicherweise vornehmlich die tatsächlichen Täter*innen – nicht oder falsch antworten und sich die übrigen dem Vorwurf ausgesetzt sehen, potentielle Täter*innen zu sein?

In der empirischen Korruptionsforschung erfreut sich trotz dieser Schwierigkeiten ein Erhebungsparadigma wachsender Beliebtheit: der vignettenbasierte Factorial Survey ²⁴ (Auspurg & Hinz, 2015). Er verbindet eine Erhebung mit den Vorteilen eines Experiments, indem den Befragten kurze Situationsbeschreibungen (sogn. Vignetten) vorgegeben werden, deren zentrale Aussagen (experimentell) verändert und zufällig den Antwortenden zugewiesen werden.

Der vorliegende Beitrag beschreibt zunächst, was die Besonderheiten der Erhebungssituation von Korruptionsdelikten sind und auch, welche Wege die bisherige Forschung gegangen ist, um Korruption auf der Individualebene zu erfassen. Im Anschluss werden die Merkmale und die Besonderheiten eines vignettenbasierten Factorial Surveys geschildert. Dabei werden die im „Gedankenexperiment“ aufgeworfenen Fragen mit den Antworten aus der bisherigen Forschung in Verbindung gebracht, die sich vor allem als „best practice“ Vorgehensweisen ergeben haben. Im Anschluss werden einige Studien dargestellt, bei denen eine Erhebung bestimmter Korruptionsdelikte wissenschaftlich verwertbar ist. Das abschließende Kapitel zeigt nochmals einige Vorteile und Schwächen, aber auch mögliches Forschungspotential mit dem Verfahren des vignettenbasierten Factorial Surveys auf.

Besonderheiten der Erfassung von Korruptionsdelikten

Korruption gilt als ein sensibles Thema, wenn es in Befragungen angesprochen wird (Dunkelmann, 2022, S. 171; Agerberg, 2022, S. 674). Ein naheliegender Grund dafür könnte sein, dass das Thema in westlich geprägten Ländern normativ negativ besetzt ist, also als etwas Verwerfliches angesehen wird. Auch wenn Befragte nicht genau die Straftatbestände der Korruption aus dem Gesetzbuch kennen, so reicht die Information aus, dass Korruption unter Strafe steht, damit Befragte ihre Antwort überdenken und möglicherweise Antworten geben, die nicht ihren eigenen Taten oder Einstellungen entsprechen. Bezogen auf die Sensitivität von Korruption stellen Delios et al. (2024, S. 238, eigene Übersetzung) gerade bei Befragungen mit Auskünften über das eigene Verhalten oder die eigene Einstellung einen systematischen Erhebungsfehler fest: *„Gesetzliche Sanktionen gegen Korruption und*

²⁴ Wir betrachten in diesem Beitrag vor allem vignettenbasierte Factorial Survey Studien. Es gibt neben diesen auch eine Reihe von ähnlichen Studienarten z.B. einfache Vignettenstudien (wie von Psychologen oftmals eingesetzt).

kulturelle Normen, die Bestechung entgegenstehen, werden Befragte dazu bringen, zu wenig zu berichten oder rundheraus zu lügen, was dazu führt, dass Berichte über das (eigene) Korruptionsverhalten in ihrer Häufigkeit nach unten verzerrt sind.“

Nach Tourangeau und Yan (2007, S. 860) lassen sich drei Dimensionen der Sensitivität im Kontext von Umfragen unterscheiden. Die erste Dimension bezieht sich auf Themen, mit denen Befragte soziale Tabus verbinden. Dies kann für Fragen nach Korruption durchaus zutreffen. Werden diese Themen in Befragungen angesprochen, dann können diese Frage als aufdringlich empfunden werden (wie dies möglicherweise beim „Gedankenexperiment“ zu Beginn des Beitrages für Sie der Fall war). Über derartige Tabuthemen wird im Alltag nicht gesprochen und entsprechend empfinden Befragte es als unangemessen, wenn Fragen danach gestellt werden. Nach Tourangeau and Yan (2007, S. 860, eigene Übersetzung) werden Befragte solche Fragen *"als Eingriff in die Privatsphäre ansehen, unabhängig davon, was die richtige Antwort für den Befragten ist."*

Die zweite Dimension der Sensitivität bezieht sich auf Fragen, bei denen die Offenlegung bestimmter Antworten eine Bedrohung darstellt. Solche Bedrohungen betreffen Antworten, die negative Konsequenzen nach sich ziehen können, z.B. wenn die Angaben der Befragten einer dritten Partei, wie anderen Personen oder Behörden, bekannt werden. Fragen nach dem (früheren) Korruptionsverhalten von Befragten sind beispielsweise aus der Sicht von Täter*innen gefährlich. Täter*innen haben kein Interesse daran, dass ihre Taten bekannt werden. Da Befragte ohne Korruptionstaten nichts zu verbergen haben, besteht bei ihnen keine Gefahr der Offenlegung. Dieses Beispiel verdeutlicht, dass diese Dimension der Sensitivität zum verzerrten Erfassen von korruptem Verhalten in einer direkten Personenbefragung führen kann, da Befragte, die Korruptionstaten in der Vergangenheit begangen haben, andere Anreize haben, ehrlich zu antworten, als Befragte, die keine Korruptionstaten in der Vergangenheit begangen haben.

Die letzte der drei Dimensionen der Sensitivität nach Tourangeau und Yan (2007) bezieht sich auf sozial erwünschtes Antwortverhalten. Darunter versteht man, wenn Befragte dazu neigen, in Übereinstimmung mit den sozialen Normen ihrer Gesellschaft oder bestimmten Gruppen zu antworten (Tourangeau et al., 2000). Entsprechend müssen nach Tourangeau und Yan (2007, S. 860, eigene Übersetzung) bestimmte Voraussetzungen erfüllt sein: *"Diese Art von Sensitivität setzt voraus, dass es klare soziale Normen für ein bestimmtes Verhalten oder eine*

bestimmte Einstellung gibt; Antworten, die über normenkonforme Verhaltensweisen oder Einstellungen berichten, gelten als sozial erwünscht und Antworten, die über Abweichungen von den Normen berichten als sozial unerwünscht." Auch bezüglich dieser Dimension kann es zu verzerrten Erfassungen kommen, wenn ein unerwünschtes Verhalten vorliegt oder eine Einstellung besteht, die gegen die soziale Norm verstößt. (Norm-)Konforme Befragte haben keinen Anreiz, ihre Antwort zu ändern. In westlichen Staaten gibt es auch wegen des strafrechtlichen Charakters korrupter Handlungen einen hohen Normdruck, sich nicht als korrupt darzustellen. Das mag umso mehr gelten, da Korruption vornehmlich im beruflichen Umfeld auftritt. Entsprechend würden Personen, die korrupte Taten begangen haben, auch negativ geächtete Aussagen über ihre Arbeit machen. Die letzten Überlegungen legen nahe, dass soziale Erwünschtheit einerseits als eine Antwortstrategie betrachtet werden kann. Andererseits wird soziale Erwünschtheit bereits seit frühen Studien zu diesem Thema als eine individuelle Antworttendenz (im Sinne eines stabilen Charaktermerkmals) betrachtet und analysiert (z.B. DeMaio, 1984). Unabhängig davon, wie soziale Erwünschtheit betrachtet wird, ist es unwahrscheinlich, dass Korruption als ein neutrales oder sogar sozial erwünschtes Verhalten in westlichen Ländern eingeschätzt wird. Das liegt auch daran, dass in Korruptionssituationen ein relativ voraussetzungsvoller illegaler Austausch stattfindet, der neben seinem gesellschaftlichen und ökonomischen Schaden auch moralisch als falsch eingeschätzt wird (Persson et al., 2013, S. 455f).

Dieser voraussetzungsvolle Austausch wird im deutschen Strafrecht damit in Verbindung gebracht, dass Korruption Freiwilligkeit zur Tat und Konsens unter den Tätern voraussetzt (Niehaus, 2019). Für die Sozialwissenschaften wird dieser voraussetzungsvolle Austausch relevant, weil er eine spezielle Beziehung unter den Täter*innen voraussetzt, die so eben nicht bei anderen Straftaten auftritt. Würde ein*e Korruptionspartner*in unfreiwillig zu einem Korruptionsaustausch gedrängt, läge eine Erpressung oder Nötigung vor, die andere soziale Interessen und Anreize nahelegen. Wer erpresst oder genötigt wird, kann sich an die Polizei oder die Strafverfolgungsbehörden wenden, eine Anzeige erstatten und einen Strafverfolgungsprozess in die Wege leiten. Typischerweise tun dies Korruptionspartner*innen nicht: sie zeigen sich nach einer erfolgreichen Korruptionstat nicht gegenseitig an, denn das würde ihre Tat öffentlich und den jeweils eigenen Korruptionsvorteil nichtig machen. Korruption setzt auf der sozialen Ebene voraus, dass sich die Beteiligten darüber einig sind, worin ihre Korruptionstat besteht und dass diese später nicht an die

Öffentlichkeit dringen darf. Es liegt eine „Unrechtsvereinbarung“ vor (Niehaus, 2019, S. 228), zu der man nicht gedrängt werden kann.

Ein Schalterbeamter kann ein Bestechungsangebot in Form eines Geldbetrages durchaus ablehnen und darauf verzichten, einen Führerschein auszustellen. Er würde eine Annahme einer Bestechung vermeiden, wenn er wüsste, dass der Bestechende später die Korruption bekannt machen würde. Umgekehrt würde ein Bestechender kein Geldangebot machen, wenn später die Korruption bekannt und ihm der Führerschein dann wieder entzogen würde.

Es gibt im deutschen Strafrecht mehrere Paragraphen, die mit Korruptionsdelikten in Verbindung gebracht werden. Diese sind Befragten in der Regel nicht bekannt. Zudem gibt es Grauzonen, bei denen die Eindeutigkeit der Zuordnung insbesondere für Laien fehlt. Was also in direkten Erhebungsfragen oder vorgegebenen Vignetten als Korruption bezeichnet wird, hat einen Einfluss auf die Überlegens- und Antwortprozesse der Befragten. In empirischen Studien sowohl bei direkten Erhebungen oder etwa bei experimentellen Spielen dominiert das Bestechungsparadigma als Abbildung von Korruption. Auf diese Weise werden keine besondere Unterformen oder Grauzonenbeispiele von korrupten Taten untersucht. Vielmehr spiegeln die untersuchten Bestechungsdelikte eindeutig Korruptionsinhalte wider, die auch in der breiten Öffentlichkeit als Korruption verstanden werden. Interessanterweise gilt das nicht nur für Deutschland, sondern auch für die meisten anderen Länder, selbst wenn dort ein hohes Korruptionsniveau vorherrscht (Fisman & Golden, 2017, S. 26ff).

In empirischen Umsetzungen zur Korruptionsthematik finden sich in der Literatur mehrere Facetten von dem, was genau untersucht und ggf. auch erklärt werden soll. Dazu gehören vornehmlich, ob bei Proband*innen eine Handlungsintention besteht, zukünftig korrupte Taten zu begehen (*Handlungsintention*), ob diese bereits in der Vergangenheit (z.B. in den letzten drei Jahren) Korruptionsdelikte begangen haben (*Prävalenz*), wie ihre Bewertung und Einstellung zu Korruptionstaten ist (*Einstellung*) oder wie sie (die Häufigkeit und Art von) Korruptionstaten anderer, etwa in ihrer Firma oder Behörde, einschätzen (*Fremdbeobachtung*).

Oliveros und Gingerich (2020, S. 384) unterscheiden zwei Gründe, warum es gerade in Erhebungen zur Korruption zu einem bewusst falschen Antwortverhalten kommen kann: einerseits sind dafür Faktoren verantwortlich, die mit der Erhebung selbst zu tun haben (wie

z.B. eine mangelnde Anonymität bei der Befragung), andererseits gibt es Faktoren, die in der antwortenden Person selbst liegen (wie etwa die Furcht vor Entdeckung).

Ein Teil der Erhebungsforschung hat sich in den letzten Jahrzehnten mit der Frage beschäftigt, wie die Anonymität der Befragten sichergestellt werden kann, so dass diese eine „wahre“ Antwort geben können, ohne später mit negativen Konsequenzen rechnen zu müssen. Es wurden eine Reihe von Techniken entwickelt, welche die Anonymität sicherstellen sollen. Dazu gehören z.B. die *Randomized Response Technik* (z.B. Krumpal & Voss, 2020) oder *Non-Randomized-Response Techniken* etwa in Form des Crosswise Models (z.B. Oliveros & Gingerich, 2020, S. 388). Allerdings setzt die Anwendung dieser Techniken eine Reihe von Annahmen voraus (wie z.B. dass die Befragten den Untersuchungsdurchführenden vertrauen, dass die Techniken auch einwandfrei angewendet werden), so dass in der Praxis nicht immer die gewünschten Ergebnisse erreicht werden, wenn diese Annahmen nicht zutreffen (Lensvelt-Mulders & Boeije, 2007; Boeije & Lensvelt-Mulders, 2002, S. 31).

Wenn Korruption mit einem vignettenbasierten Factorial Survey untersucht wird, wird versucht, die Sensitivität über den Präsentationsstil und über die indirekte Fragenart abzumildern (Sauer et al., 2020, S. 197). Wie dies geschieht, wird im nächsten Kapitel näher dargestellt.

Der vignettenbasierte Factorial Survey

Ein vignettenbasierter Factorial Survey kombiniert in einer Befragung die Lebensnähe und kontextgebende Eigenschaft von Vignetten mit einem experimentellen Design (Auspurg & Hinz, 2015). Vignetten sind kurze, fiktionale Beschreibungen von Situationen oder Personen. Im Rahmen des experimentellen Designs werden unterschiedliche Ausprägungen dieser Variablen (sogn. Level) zufällig in der Befragung variiert. Verschiedenen Befragten werden somit Vignetten präsentiert, die jeweils andere Ausprägungen der Variablen enthalten. Die Befragten sollen dann Antworten zu diesen fiktionalen Beschreibungen geben.

Vignettenbasierte Factorial Surveys gelten in Forschungssituationen als nützlich, in denen Beobachtungen des sozialen Phänomens oder Experimente unter Realbedingungen nicht möglich oder unethisch wären (Auspurg & Hinz, 2015). Korruptionssituationen fallen in diese Kategorie. Zum jetzigen Zeitpunkt lassen sich Korruptionsdeals zwischen den Täter*innen

weder beobachten, noch können Bestechungen mit Konsequenzen in der Realität (z.B. mit einer entsprechenden Strafverfolgung) experimentell durchgeführt werden. In experimentellen Spielen, in denen die Teilnehmenden Korruptionssituationen nachstellen und mit echten Gewinnen oder Verlusten (z.B. an Spielgeld) umgehen (z.B. Abbink, 2002 oder Banerjee & Mitra, 2018), enden die Konsequenzen für korrupte Spielpartner*innen mit dem Ablauf des Spiels. Trotzdem können experimentelle Spiele die Mechanismen erforschen, die bei Entscheidungen über Korruption eine Rolle spielen.

Vignettenbasierte Factorial Surveys können wegen der experimentellen Variation der erklärenden Variablen in gleicher Weise Mechanismen der Korruption erforschen (Graeff, 2016). In Bezug auf die Erforschung des Auftretens oder des Bekämpfens von Korruption, enthalten die Vignetten die Variablen, von denen aufgrund theoretischer oder empirischer Vorgaben erwartet werden kann, dass sie das Auftreten oder das Bekämpfen von Korruption beeinflussen. Dies kann beispielsweise die Höhe des Geldvorteils für eine*n Bestochene*n sein oder dessen Wahrscheinlichkeit, bei der Korruption entdeckt zu werden. Die Szenarien setzen die Antworten der Befragten in einen verständlichen Kontext, wenn die Vignetten die Inhalte realitätsnah abbilden („real-life-Stories“, Sauer et al., 2020, S. 197). Es können mehrere Erklärungen bzw. Ursachen für Korruption gleichzeitig untersucht und damit dem multidimensionalen Phänomen der Korruption Rechnung getragen werden. Vorteilhaft ist auch, dass die Befragten ein ähnliches Grundszenario erhalten, der allgemeine Kontext ihrer Antworten also gleichbleibt.

In erhebungstechnischer Hinsicht führt die Kombination aus experimentellem Design und Erhebung dazu, dass große Fallzahlen (vor allem auf der Ebene der Vignetten) realisiert werden können (siehe Aguinis & Bradley, 2014) und sich damit die Teststärke erhöht. Außerdem lassen sich u.U. Generalisierungen der Untersuchungsergebnisse ableiten, was üblicherweise mit rein experimentellen Forschungsstudien in den Sozialwissenschaften selten möglich ist (Mutz, 2011, S. 22). Durch das experimentelle Design kommen zusätzliche Vorteile dazu, die bei einfachen Befragungen so nicht existieren. Dazu gehört, dass die Zufallszuordnungen der erklärenden Variablenausprägungen diese Variablen unabhängig voneinander machen und ihre einzelnen Einflüsse damit sichtbar gemacht werden können.

Wie bereits angedeutet, gelten Vignetten als eine Möglichkeit, verschiedene Arten der Sensitivität bei Erhebungen zu reduzieren (wie z.B. der sozialen Erwünschtheit, Auspurg et al.,

2014; Wallander, 2009). So stellen beispielsweise Armacost et al. (1991) in einer Briefbefragung von Firmenchefs von Produktionsbetrieben dieselben sensitive Fragen direkt, über Vignetten oder über die Randomized-Response Methode. In diesen Fragen ging es um fragwürdiges wirtschaftliches Verhalten, wie das Ausnutzen von Insider-Informationen oder die Durchführung von vorgetäuschten Job Interviews, das entweder als Selbstreport oder der Wahrnehmung des Verhaltens anderer erhoben wurde (mit Ausnahme der Randomized-Response-Befragung). Das direkte Fragen provozierte beim Selbstreport mehr sozial erwünschte und gleichzeitig weniger sozial unerwünschte Antworten als die beiden anderen Methoden. Die Generalisierbarkeit der Ergebnisse wird wegen der geringen Antwortrate von 20,2% von den Autoren selbst in Frage gestellt (Armacost et al., 1991, S. 1084).

Ein Selektionseffekt derjenigen, die solche Vorgehensweisen typischerweise verwenden, kann sich in der hohen Unit-Nonresponse-Rate widerspiegeln. So argumentieren etwa Jensen et al. (2010), dass bei Befragungen zur Korruptionswahrnehmung von Firmenangehörigen Nonresponse und verzerrte Antworten einen selbstschützenden Mechanismus darstellen, der vom politischen Umfeld abhängt. Sie zeigen anhand von Weltbankdaten zu über 44.000 Firmen in 72 Ländern, dass in politisch repressiven Ländern häufiger Nonresponse und verzerrte Antworten bei Befragungen zu Korruptionsinhalten vermutet werden können.²⁵

Weitere Gründe, warum vignettenbasierte Factorial Surveys bessere Ergebnisse gegenüber einer direkten Befragung bezüglich sensibler Inhalte erbringen sollen, sind die hypothetischen Beschreibungen und die üblicherweise mehrfachen Erklärungsvariablen (Auspurg et al., 2014, S. 141). Wenn ein*e Protagonist*in in der Vignette beschrieben wird und die Befragten an seiner/ihrer Stelle antworten sollen, dann geben sie damit lediglich indirekt ihre eigene Meinung zu dieser hypothetischen Situation zum Ausdruck. Da meistens die Levels mehrerer Erklärungsvariablen über die verschiedenen Vignetten variiert werden, wird die Aufmerksamkeit für eine einzelne (möglicherweise besonders sensitive) Variable bei der Beantwortung reduziert (Wallander, 2009, S. 506).

²⁵ In dieser Studie werden mögliche Antwortverzerrungen (*potencial false responses*) aus der Diskrepanz zwischen den Befragungsdaten und der Experteneinschätzung des Ausmaßes der Korruption in den Ländern abgeleitet. Allerdings wird in der Studie nicht nach dem Ausmaß der Korruption gefragt, sondern danach, ob Korruption ein Problem darstellt und inwieweit die Schwere des Problems als Hindernis eingestuft wird (5-stufige Skala). Wer sich mit Korruption als einer notwendigen Praktik arrangiert hat, dürfte hier kein Problem sehen, was den Vergleich zur Experteneinschätzung in der Aussagekraft reduziert (Jensen et al., 2010, S. 1485).

In der Studie von Auspurg et al. (2014), in welcher die Befragten die Fairness von Verdienstzahlungen in Abhängigkeit vom Geschlecht einschätzten, fand sich ebenfalls Evidenz dafür, dass vignettenbasierte Factorial Surveys die Wirkung sensibler Inhalte reduzieren können – verglichen zu der Situation, in der Erhebungsteilnehmende direkt befragt werden (Auspurg et al., 2014, S. 143f). In ihrer Untersuchung fanden sich darüber hinaus auch Hinweise dafür, dass der Effekt der Sensitivität insbesondere für höher gebildete Personen zurückgeht (Auspurg et al., 2014, S. 145).

Korruptionsstudien

Der vignettenbasierte Factorial Survey ist ein Verfahren, um theoretische Vorgaben oder Hypothesen an Individuen zu untersuchen. Es setzt eine Theorie oder zumindest eine begründete Hypothese voraus, um die Erklärungsvariablen (manchmal auch Dimensionen genannt) und ihre Ausprägungen (Levels) für die Vignetten auszuwählen und zu designen (Auspurg & Hinz, 2015). Es gibt in der Korruptionsforschung bisher keine einheitliche Theorie, welche die Entstehung oder Aufrechterhaltung von Korruptionspraktiken auf Individualebene umfassend erklären und als Grundlage für das inhaltliche Design eines vignettenbasierten Factorial Surveys dienen könnte. Forschende rekrutieren ihre theoretischen Vorgaben entsprechend aus sozialwissenschaftlichen Disziplinen (wie der Politikwissenschaft, Psychologie oder Soziologie) oder aus deren Nachbarwissenschaften wie der Rechtswissenschaft oder der Ökonomie.

Dies wird auch in den folgenden Studien deutlich, die beispielhaft ausgesucht wurden, um die Anwendung vignettenbasierter Factorial Surveys in der Korruptionsforschung zu illustrieren.

Politikwissenschaftliche Studien zielen zumeist auf die Erforschung politischer Korruption ab, wobei disziplintypische Unterschiede in der Anwendung von Factorial Surveys deutlich werden. So variiert Vera (2020) drei Aspekte in der Beschreibung eines fiktiven, peruanischen Kongresskandidaten. Im Fokus stehen dabei, ob er bereits wegen Korruptionsvergehen beschuldigt worden war oder nicht, ob er mehr Erfolge für die lokale Gemeinschaft erreicht hat als die Mehrheit der Bürgermeister (oder nicht) und ob er sich in einem Wahlkreis bewirbt, in dem das Niveau der Korruption hoch ist oder nicht. Die Vignetten wurden als Teil einer landesweiten Erhebung in Peru eingesetzt. Die erklärte Variable misst (als

Handlungsentention) die Einschätzung der 1308 peruanischen Befragten, wie wahrscheinlich es ist, dass diese für den jeweils beschriebenen Kandidaten stimmen werden.²⁶ Diese Einschätzung sollten die Befragten auf einer 7-stufigen Skala angeben. Die Ergebnisse legen einen negativen Einfluss der Beschuldigung von Korruptionsvergehen auf die Wahrscheinlichkeit nahe, für den Kandidaten zu stimmen. Aufgrund des multifaktoriellen Designs konnten die Wechselwirkungen mit anderen Dimensionen betrachtet werden. Dabei zeigt sich, dass die politischen Kandidaten weniger für Korruptionsvergehen abgestraft werden, die sich als kompetent erwiesen haben.

Aufbauend auf emotionstheoretischen Überlegungen der politischen Psychologie und mit einem ähnlichen Design wie Vera (2020) untersuchen Birch et al. (2017, S. 901), ob bestimmte Verhaltensweisen, die in den Grauzonenbereich von Korruption fallen, zu einer negativen Bewertung bei Befragten führen. Die Erhebung wurde in Frankreich durchgeführt und stellt im Gegensatz zu Vera (2020) somit ein Beispiel für eine Untersuchung in einem westlichen Land dar. Die zu bewertenden Verhaltensweisen waren auf den französischen Kontext ausgerichtet und enthielten eine Doppel- vs. Einfachmandatierung, einen möglicherweise illegitimen politischen Vorteil (oder die bewusste Betonung, dass dies nicht der Fall sei) und einen mit Selbstinteressen verbundenen möglichen finanziellen Vorteil (oder die Abwesenheit einer Verquickung mit eigenen Vorteilen). Die abhängige Variable misst auf einer 11-er Skala, ob die Verhaltensweisen aus Sicht der Befragten akzeptierbar sind. Die Ergebnisse zeigen, dass illegitime politische Vorteile und selbstinteressenbezogene finanzielle Vorteile signifikant weniger als akzeptabel eingeschätzt werden. Weiterhin bringen Birch et al. (2017, S. 905) diese Ergebnisse mit Klassifikationen der Befragten nach emotionspsychologischen Zuordnungen von Angst und Wut in Verbindung, wobei sich die daraus ergebenden Resultate als nicht konsistent erweisen.

²⁶ Eine Besonderheit in der Anwendung des Factorial Surveys zeigt sich in der Studie von Vera (2020) im experimentellen Design. Jeder befragten Person wurde genau eine Vignette zufällig zugeordnet (sogn. „Between-subject design“), was von Vera (2020, S. 660, eigene Übersetzung) in folgender Weise begründet wird: „[...] dieser experimentelle Ansatz wurde als unaufdringlicher Weg gewählt, um die sozial erwünschte Einstellung der Ablehnung eines korrupten Kandidaten zu messen. Erstens müssen die Befragten nicht mehrere Vignetten beantworten, was sie auf die experimentellen Variationen aufmerksam machen würde. Zweitens wird ihnen eine Vignette mit einem recht komplexen hypothetischen Kandidaten vorgelegt, der mehrere Eigenschaften aufweist, was es dem Befragten erschwert, die in der Vignette manipulierten Schlüsselfaktoren zu erkennen. Und schließlich, wird nach der Einstellung einer dritten Person gefragt, so dass der Befragte ehrlich antworten kann, ohne die eigene Präferenz explizit zu nennen.“

Soziologische Forschungen, die Korruption mit vignettenbasierten Factorial Surveys untersuchen, greifen auf andere theoretische Ausgangspunkte zurück, wie dass Korruption eine Entscheidung zwischen einer partikularen und universellen Norm enthält (Schweitzer, 2022) oder eine Entscheidung darstellt, die von konkreten Anreizen motiviert ist (Graeff, 2016). Entsprechend zielen die Vignetten und erklärten Variablen auf Norm- oder Werturteile ab (Auspurg & Hinz, 2015) oder auf eine Intention der Befragten, an der Stelle des in der Vignette beschriebenen Charakters zu entscheiden. Aus Sicht von Liebe et al. (2020, S. 163) werden solche Norm- oder Werturteile in geradezu eleganter Weise gemessen, weil die zugrundeliegenden (theoretischen) Konzepte und Inhalte in ihrer Bedeutung zu den korrespondierenden Situationsvariablen in den Vignetten erfasst werden und nicht (wie bei Direktbefragungen oftmals) als Einzelitem abgefragt werden.

Ein Beispiel für gruppenbezogene Werturteile bezüglich korrupter Handlungen stellt die Studie von Kleinewiese und Graeff (2020) dar. An der Zielgruppe der Freiwilligen Feuerwehr wurde untersucht, ob der Teamgeist (als eine Form von Gruppennorm) dazu führen kann, dass abweichende Verhaltensweisen von Teammitgliedern einen Konflikt zwischen partikularen und universalistischen Normen auslösen können. Die Arbeit bei der Freiwilligen Feuerwehr ist ähnlich wie die bei der Polizei und dem Militär stark davon abhängig, dass ein Teamzusammenhalt existiert, damit Herausforderungen für oder Angriffe auf das Team gemeinsam gemeistert werden können. Dieser Teamgeist legt aber auch eine Schweigenorm („code of silence“) nahe, wenn ein Teammitglied für deviante Verhaltensweisen eigentlich gemeldet werden sollte. Umgekehrt legt die Beobachtung eines juristisch ahndungsbedürftigen Verhaltens eines Teammitglieds das Anzeigen der Tat nahe, so wie dies von universalistischen Normen bzw. den Gesetzen gefordert wird. Aus Sicht früherer Studien existiert hier ein Konflikt zwischen der Loyalität zum Team und von Gesetzen gegebenen Fairnessnormen, der in der Korruptionsforschung dem „Whistleblowing“ zugeschrieben wird (Dungan et al., 2019). Die abhängige Variable war eine Prozentangabe darüber, für wie wahrscheinlich die Mitglieder der Freiwilligen Feuerwehr es halten, dass die in den Vignetten beschriebene Person das deviante Verhalten eines Teammitglieds der Freiwilligen Feuerwehr dem Vorgesetzten meldet. Gemessen wurde auf einer 11-stufigen Skala. Die Ergebnisse zeigen, dass ein starker Teamgeist dazu führen kann, dass abweichende Verhaltensweisen von Teammitgliedern den zuvor genannten Konflikt auslösen. Wenn in solchen Situationen Angriffe von außerhalb auf das Team einwirken, verstärken diese Angriffe den „code of

silence“. In den Vignetten wurde die Tötlichkeit von betrunkenen Fußballfans als Angriffe auf bzw. Störung von einem Einsatz der Freiwilligen Feuerwehr variiert. Da es keine originäre Korruptionstheorie gibt, die als Ausgangspunkt für solche Untersuchungen dienen kann, wurde in diesem Fall versucht, die Resultate in einem breiteren devianztheoretischen Rahmen einzuordnen. Andere Studien zu Devianz und Kriminalität gehen ähnlich vor und greifen theoretische Ausgangspunkte der *General Theory of Crime* oder der *Situational Action Theory* (z. B. Kleinewiese, 2022) auf.

In einer Untersuchung von Graeff et al. (2014) wurden Studierende dazu aufgefordert, basierend auf Szenarienbeschreibungen zu entscheiden, ob sie Bestechungsgelder annehmen würden. Die Teilnehmenden wurden durch die Vignettenbeschreibungen in die Rolle eines Tutors versetzt. Der Tutor konnte darüber entscheiden, ob er einer Studierenden das Bestehen einer Prüfung bescheinigen sollte, obwohl sie diese nicht bestanden hatte. In den Vignetten wurden Entscheidungsfaktoren, wie die Bestechungssumme, potenzielle Strafen, die Wahrscheinlichkeit entdeckt zu werden und die Vertrauenswürdigkeit der Studierenden (und somit die Erfolgswahrscheinlichkeit der korrupten Handlung) variiert. Die Befragten konnten nach jeder Vignette auf einer 10-Punkte Skala angeben, ob sie (im Sinne einer Verhaltensintention) bereit wären, die Note der Studierenden zu ändern. Die Befunde zeigen, dass mit steigenden Anreizen (wie höheren Bestechungssummen oder größerer Vertrauenswürdigkeit) auch die Bereitschaft zu korruptem Verhalten zunimmt. Diese Bereitschaft sinkt mit größerer Entdeckungswahrscheinlichkeit. Bemerkenswert war das Ergebnis, dass normative Entscheidungsfaktoren einen größeren Erklärungswert besitzen und mehr Varianz aufdecken als die zuvor genannten Anreize bzw. Abschreckungsfaktoren. Dies deutet darauf hin, dass auf praktischer Ebene zur Bekämpfung von Korruption das Fördern normativer Entscheidungselemente (wie der Bedeutungserhöhung von Fairnessnormen) effektiver sein könnten als einfache Belohnungs-Straf-Mechanismen (z.B. Graeff, 2016).

Psychologische Studien, die Korruption untersuchen, verwenden meist Traits als erklärende Variablen und Verhaltensintentionen (für korrupte Handlungen) als erklärte Variable. Sie gestalten die Vignetten hinsichtlich eines unethischen Entscheidungsverhalten bzw. als ethische Dilemma-Situation (Rabl, 2019, S. 79). Ähnlich wie in der Studie von Graeff et al.

(2014) betrachten De Waele et al. (2021) die Bestechungswilligkeit von Studierenden.²⁷ Die Autor*innen verwenden eine Vignettenstudie in Form eines s.g. between-subject Designs in Belgien, Deutschland sowie den Niederlanden. Ein solches Design liegt vor, da jede*r Studierende nur eine (von drei Variationen der) Vignette erhält. Die Studierenden wurden gebeten sich vorzustellen, dass sie wieder im ersten Semester ihres Studiums seien und eine Prüfung nicht bestanden hätten. Weiterhin hätten die Studierenden die Rückmeldung erhalten, dass sie 9,4 von 20 Punkten erreicht hatten, wobei bereits 9,5 Punkte zum Aufrunden der Punktzahl und somit zum Bestehen der Prüfung geführt hätten. In Anschluss an diese Einführung wird dann eine von drei Vignetten zufällig angezeigt, welche folgende Stufen von Bestechung beschreibt:

- 1) Emotionale Bitte an den Dozierenden, die Prüfung als Bestanden einzutragen (von den Autor*innen als *white bribery* bezeichnet).
- 2) Vorschlag eines Gefallens an den Dozierenden (Reparatur dessen defekten Autos) im Gegenzug für das Bestehen der Prüfung (von den Autor*innen als *grey bribery* bezeichnet).
- 3) Übergabe eines Umschlags mit 500€ an den Dozierenden verbunden mit der Bitte die Note zu überdenken (von den Autor*innen als *black bribery* bezeichnet).

In allen drei Fällen enthält die Vignette zudem den Hinweis, dass aus Gerüchten bereits vermutet werden könne, dass diese Verhaltensweise jeweils erfolgreich zum Bestehen der Prüfung führe. Als eine von zwei zentralen erklärenden Variablen verwenden De Waele et al. (2021, S. 5, eigene Übersetzung) die Social Value Orientation (SVO): „Dieses Konzept bezieht sich nach Messick und McClintock (1968) auf eine stabile Präferenz für bestimmte Ergebnisse für sich selbst gegenüber den Ergebnissen anderer und erfasst damit das Ausmaß, in dem ein Individuum intrinsisch hauptsächlich um das persönliche Wohl gegenüber dem der Gruppe besorgt ist.“ Während für die zweite Erklärungsvariable (Public Service Motivation) kein

²⁷ Abweichend von anderen Studien wurde die abhängige Variable der Bestechungswilligkeit (*willingness to bribe*) nicht als direkte Frage erhoben, sondern als geometrischer Mittelwert aus vier Variablen gebildet (De Waele et al., 2021, S. 5): Für wie wahrscheinlich es die Befragten halten, in dem in der Vignette beschriebenen Kontext zu bestechen (*likelihood*), für wie gerechtfertigt sie die Bestechung halten (*justification*), wie wohl sie sich mit der Bestechung fühlen (*affect*) und ob das Bestechen ein Fehler sein würde (*mistake*).

bedeutsamer Erklärungsbeitrag gefunden werden kann, erhöht (aus Sicht der Autor*innen paradoxer Weise) SVO die Bereitschaft, eine Korruption zu begehen.²⁸

In ökonomischen Studien, die auf vignettenbasierte Factorial Surveys zur Erklärung von Korruption zurückgreifen, stehen sowohl Anreize bzw. Kosten als auch personenbezogene Variablen im Vordergrund. Beispielhaft sei hier die Studie von Dickel und Graeff (2018) genannt, wo Entrepreneur*innen in einem für sie realistischen Szenario danach befragt wurden, ob sie einen ausländischen Beamten bestechen würden, um mit einer innovativen Technologie die Luft in einer Großstadt zu verbessern. Auch hier war die Handlungsintention die erklärte Variable. Die Stichprobenteilnehmenden geben einen signifikanten Einfluss der Gewinnerwartung und der Kosten sowie der Fähigkeit an, dass die Korruption erfolgreich durchgeführt werden kann. Diese Einflussgrößen werden durch die persönliche Tendenz verstärkt, über die Vor- und Nachteile der Gelegenheit für eine Korruption nachzudenken (Deliberation). Entgegen der theoretischen Erwartung (aufgrund gängiger Kriminalitätstheorien), dass die Entdeckungswahrscheinlichkeit gegen das Durchführen einer Straftat wirkt, erwies sich die Entdeckungswahrscheinlichkeit in dieser Studie als nicht statistisch bedeutsam.

Diskussion und Ausblick

Zu Beginn dieses Beitrags haben wir Sie darum gebeten sich in ein Gedankenexperiment hineinzusetzen. Sie sollten sich vorstellen, von einem Befragungsinstitut angerufen zu werden und gefragt zu werden, ob Sie in den letzten fünf Jahren eine Korruption begangen haben.

Wenn Sie sich beim ersten Lesen dieser Aufforderung (in der Einleitung dieses Kapitels) in diese fiktive Situationsbeschreibung hineinversetzt haben, dann haben Sie das getan, was von Befragten gefordert wird, wenn Sie mit einem Vignettenszenario konfrontiert werden.

Die Art der Befragung (telefonische Befragung) sowie der Zeitraum, für welchen Sie die Frage beantworten sollten (5 Jahre), hätten in einem vignettenbasierten Factorial Survey

²⁸ Gegeben die empirischen Befunde, bei denen Korruption als Diskrepanz zwischen Gruppenloyalität und Fairnessnormen betrachtet wird (Kleinewiese & Graeff, 2020; Dungan et al., 2019; Schweitzer, 2022), zeigt dieses Ergebnis, dass die Tendenz zur Gruppenloyalität Bestechung begünstigt.

Dimensionen sein können, die unterschiedliche Levels hätten aufweisen können. So hätte in einer anderen Vignette etwa nach dem letzten Jahr oder den letzten drei Jahren gefragt werden können. Im Anschluss an die Vignette wird mindestens eine Frage mit diesem Szenario verknüpft. Wenn in der Vignette bzw. dem Szenario eine Person beschrieben wird, an deren Stelle man eine Entscheidung fällen soll, dann würde man als Befragter nicht direkt auf die Frage antworten, sondern mit einer Distanz zur eigenen Person.

Wie im vorherigen Beitrag dargelegt wurde, kann ein derartiges Design Vorteile in Bezug auf die Erhebung verschiedener Facetten von Korruption mit sich bringen. Es gibt aber auch Schwierigkeiten. Zu diesen gehört die Erstellung plausibler und aussagekräftiger Szenarien. Dies erfordert in der Regel Vorwissen über den zu untersuchenden Kontext, um relevante, erklärende Faktoren (Dimensionen), aber auch plausible Levels in die Vignetten einzubeziehen. Dieses Wissen sollte dabei aus theoretischen Überlegungen oder vorherigen Studienergebnissen abgeleitet werden. Dies gestaltet sich in der Korruptionsforschung wegen des Theoriedefizits schwierig (Graeff, 2019, S. 293). Es wurde im vorherigen Abschnitt deutlich, dass dieses Defizit erfordert, theoretische Überlegungen aus unterschiedlichen wissenschaftlichen Disziplinen hinzu zu ziehen. Damit bleiben sozialwissenschaftliche vignettenbasierte Factorial Surveys oftmals disziplinbezogen und werden von Nachbarwissenschaften wegen der begrifflichen und theoretischen Inkommensurabilität²⁹ nicht aufgegriffen.

Vor der Durchführung benötigen vignettenbasierte Factorial Surveys die Durchführung fundierter Pretests in der interessierenden Grundgesamtheit. Es muss sichergestellt werden, dass die verwendeten Dimensionen (Faktoren) und zugehörigen Level verstanden und in die Antwortfindung einbezogen werden. Dazu muss der Pretest Hinweise dafür geben, dass die Variation der erklärenden Faktoren (der experimentelle Aspekt der Befragung) das Ergebnis verändert. Ist dies nicht der Fall, müssen die Vignettendimensionen und Levels entsprechend angepasst und erneut getestet werden. Dieser Prozess ist im Vergleich zu einer „einfachen“ Befragung mit einem standardisierten Fragebogen (welcher ggf. auf bereits bewährte Items aus anderen Befragungen zurückgreifen kann) aufwändig und zeitintensiv. Zusätzlich ist für Anwendende dieses Verfahrens zu bedenken, dass die Durchführung eines vignettenbasierten

²⁹ Inkommensurabilität bedeutet, dass zwei Theorien so grundlegend unterschiedliche Annahmen oder Konzepte verwenden, dass ein direkter Vergleich oder eine Übersetzung der einen Theorie in die Begriffe der anderen schwierig bis unmöglich ist.

Factorial Surveys den Befragten im Vergleich zu standardisierten Befragungen üblicherweise eine höhere kognitive Leistung abfordert.

Den genannten Herausforderungen stehen Vorteile für die Erfassung sensibler Themen, wie Korruption eines ist (Dunkelmann, 2022, S. 171), gegenüber. Betrachtet man die drei Dimensionen der Sensitivität nach Tourangeau und Yan (2007) – gesellschaftliche Tabus, Angst vor Entdeckung und soziale Erwünschtheit – so kann eine Antwortverzerrung auf individueller Ebene trotz der Erhebung mit vignettenbasierten Factorial Surveys ebenso wenig ausgeschlossen werden wie eine Antwortverweigerung. Aus Sicht der Untersuchenden äußert sich die Antwortverzerrung beispielsweise bei sozialer Erwünschtheit darin, dass auf die Frage nach der Vignette keine Zustimmung zu korrupten Taten bzw. keine entsprechende Handlungsintention angegeben wird. Dadurch variieren die Antworten der Befragten nicht über die Ausprägungen der erklärten Variablen hinweg. Im Extremfall geben alle Befragte exakt die gleiche Antwort: es gibt keine Zustimmung zu korrupten Taten oder keine entsprechende Handlungsintention. Dann wäre die Befragung nicht auswertbar. Wenn viele Befragte von der Möglichkeit Gebrauch machen, die Frage nach der Vignette überhaupt nicht zu beantworten (Item-Nonresponse), reduziert sich die Anzahl der Antworten, die auswertbar sind. Die im vorherigen Abschnitt genannten Studienergebnisse zeigen aber, dass auch sensitive Fragen in der Regel noch hinreichende Variation der Befragtenantworten erreichen, so dass eine Auswertung möglich ist.

Das mag auch daran liegen, dass Befragte in der Regel nicht direkt auf eine Frage mit Korruptionsbezug antworten müssen, sondern vielmehr eine hypothetische Situation beurteilen. Betrachtet man unter dieser Prämisse die erste Dimension von Tourangeau und Yan (2007; gesellschaftliche Tabus), so kann die Verwendung von vignettenbasierten Factorial Surveys zu einer abgemilderten Wahrnehmung des Tabus führen. Befragte könnten sich zu einem tabuisierten Thema äußern, da sie lediglich indirekt ihre Einstellung diesbezüglich angeben müssen.

Auch in Bezug zur zweiten Dimension von Tourangeau und Yan (2007) lässt sich eine Minderung der Sensitivität aufgrund der indirekten Befragungsart ableiten. Wie die referierten Studien im vorherigen Abschnitt zeigen, werden die Befragten nach einem Szenario in der Regel nach einer Handlungsintention oder Einstellung und nicht nach vergangenen Handlungen gefragt. Weiterhin beurteilen Befragte häufig Szenarien, die

gedanklich weniger mit Ihnen persönlich zu tun haben, da beispielsweise das Verhalten von Protagonist*innen in der Vignette bewertet wird, sodass weniger Reflexion des eigenen Verhaltens stattfindet. Dies schwächt die Angst vor Entdeckung ab, da kein konkretes Vergehen zugegeben werden muss. Dieses Ausbleiben von sogn. self-disclosure (dt. in etwa: Selbst-Enthüllung oder Selbst-Veröffentlichung) schwächt damit auch die Angst vor Entdeckung, da kein konkretes Vergehen an Dritte gemeldet werden kann.

Bezüglich der sozialen Erwünschtheit – der dritten Dimension nach Tourangeau und Yan (2007) – kann auch hier in Abgrenzung zur direkten Frage von einer schwächer wirkenden Sensitivität in Bezug auf Korruption ausgegangen werden. Für andere Themenbereiche, die durch soziale Erwünschtheit geprägt sind, konnte dieser Zusammenhang bereits gezeigt werden, weshalb auch bezüglich der Handlungsintention und der Einstellungen zu Korruption von dieser Wirkweise ausgegangen werden kann (Auspurg et al., 2014, S.146). Gleichwohl kann der zweiten Betrachtungsweise von sozial erwünschtem Antwortverhalten als stabilem Persönlichkeitsmerkmal (DeMaio, 1984) – ggf. im Sinne eines beschönigten Selbstbildes – im Rahmen eines vignettenbasierten Factorial Surveys nicht Rechnung getragen werden. Wenn befragte Personen bei der Erwähnung bestimmter Dimensionsausprägungen (z.B. bei Beschreibung einer Bestechung) in Richtung der sozialen Normen antworten, anstelle der Antwort, die in Übereinstimmung mit ihrer persönlichen Einstellung oder tatsächlichen Handlungsintention wäre, so bleibt dieses Problem trotz dieser Befragungsmethode bestehen.

Fasst man die bisherigen Ausführungen zusammen, lässt sich die Schärfe aus den eingangs gestellten Fragen nehmen, in denen es um die Unterstellung von Straftaten, den Grund des ehrlichen Antwortens oder der Selbstbeschuldigung bei Korruptionsinhalten ging.³⁰ Dabei ist ein zentrales Argument die indirekte Konfrontation mit Szenarien anstelle von direkten Fragen. Anstatt Befragten mögliche Korruptionstaten zu unterstellen, wird um die Einschätzung hypothetischer Szenarien gebeten. Wird dies mit einer Frage nach

³⁰ Zusätzlich zu diesen Fragen wurde die Problematik der fehlenden Eindeutigkeit des Korruptionsbegriffs aufgeworfen. Dieser Problematik kann im Rahmen eines vignettenbasierten Factorial Surveys auf zweierlei Art begegnet werden:

1. Befragte können um eine Einschätzung gebeten werden, ob das beschriebene Szenario eine Korruption darstellt.
2. In der Vignette kann eine konkrete Korruptionshandlung beschrieben werden, die sich nicht im sogn. „Graubereich“ befindet, sondern eindeutig von Befragten als Korruption erkannt wird (z.B. die Annahme von Bestechungsgeld, Graeff et al., 2014).

Handlungsintentionen und/ oder Einstellung zu Korruption verbunden, schließt es zwar self-disclosure nicht aus (z.B. Zugeben einer Einstellung zu Korruption), die Antwort ist aber weniger konkret auf vergangene Handlungen von Befragten bezogen. Da self-disclosure in diesem Fall nicht in gleichem Ausmaß mit den negativen Konsequenzen (Angst vor Entdeckung) einhergeht, wie das Zugeben einer Korruptionstat in einer direkten Befragung, könnten Vignetten es auch für Täter*innen erleichtern ehrlich zu antworten. Mehr noch: Handlungsintentionen oder Einstellungen sind nicht strafrechtlich relevant, was juristische Implikationen oder Aspekte der Strafverfolgung ausschließt. Für die Untersuchung von kriminellen Phänomenen könnten die Resultate, die mit Hilfe von vignettenbasierten Factorial Surveys gewonnen werden, deutlich weniger durch Antwortverzerrungen von Befragten negativ beeinflusst werden und Antwortverweigerungen sollten ebenfalls seltener auftreten.

Als Befragungsmethode werden vignettenbasierte Factorial Surveys auch zukünftig in der Korruptionsforschung Verwendung finden. Ihre Vorteile überwiegen die Probleme und den Aufwand ihrer Erstellung. Besonders hervorzuheben ist, dass diese Methode in der Lage ist, die Vielzahl von Faktoren, die Korruption beeinflussen, darzustellen und vor allem in ihrem Einfluss isolieren zu können. Indem realitätsnahe Szenarien dargeboten werden, können Entscheidungen und Normeinschätzungen in lebensnahen Situationen nachgebildet werden. Damit wird in eingeschränkter Weise auch die Kluft zwischen abstrakten Bewertungen ethischer Standards und tatsächlichem Verhalten bzw. von Verhaltensintentionen in realen Situationen überwunden. Da der Beobachtung und Untersuchung von echtem, korruptem Verhalten ethische Grenzen gesetzt sind, bieten vignettenbasierte Factorial Surveys eine ethisch vertretbare Möglichkeit zu untersuchen, wie Befragte bei Korruptionsentscheidungen reagieren, ohne dass sie eigenes moralisch fragwürdiges Verhalten zeigen oder sich einem unmittelbaren Risiko aussetzen. Zusätzlich kann die Methoden an verschiedene Stichproben und Kontexte angepasst werden und zur Aufdeckung von Korruptionsmechanismen eingesetzt werden, selbst bei solchen Befragungsgruppen, die für alternative Erhebungsmethoden wie die Teilnahme an experimentellen Spielen nur bedingt motiviert werden können wie z.B. Manager (Dickel & Graeff, 2016). Wenngleich einige dieser Vorteile diese Methoden gegenüber herkömmlichen Befragungen positiv abgrenzen, zeigen einige Vorteile auch prinzipielle Schwierigkeiten im Umgang mit der Erforschung von Korruption auf. Dazu gehört, dass die hypothetischen Szenarien eben hypothetisch bleiben und dass Befragte, die sich dies bewusst machen, Abweichungen von ihrem Verhalten in realen Situationen

angeben können. Außerdem werfen die Art und die Modi der Durchführung (wozu bspw. die Szenario- und Fragendarbietung, die Anzahl der Dimensionen sowie Vignetten oder die Möglichkeit der Teilnahme an der Befragung gehören, vgl. Leitgöb & Wolbring, 2021) noch eine Reihe von Fragen auf, die sich erst in zukünftigen Studien beantworten lassen.

Literatur

- Agerberg, M. (2022). Corrupted estimates? Response Bias in Citizens Surveys on Corruption. *Political Behavior*, 44, 653-678.
- Aguinis, H. & Bradley, K.J. (2014). Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies. *Organizational Research Methods*, 17(4), 351–371.
- Armacost, R.L., Hosseini, J.C., Morris, S.A. & Rehbein, K.A. (1991). An Empirical Comparison of Direct Questioning, Scenario, and Randomized Response Methods for Obtaining Sensitive Business Information. *Decision Science*, 22, 1073-1090.
- Auspurg, K., Hinz, T., Liebig, S. and Sauer, C. (2014). The Factorial Survey as a Method for Measuring Sensitive Issues. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, & P. Sturgis (Hrsg.), *Improving Survey Methods*, 137-149.
- Auspurg, K., & Hinz, T. (2015). Factorial survey experiments, Sage, Thousand Oaks.
- Banerjee, R., & Mitra, A. (2018). On monetary and non-monetary interventions to combat corruption. *Journal of Economic Behavior and Organisation*, 149, 332-355.
- Boeije, H. & Lensvelt-Mulders, G. J. L. M. (2002). Honest by Chance: A Qualitative Interview Study to Clarify Respondents' (Non-)Compliance With Computer-Assisted Randomized Response. *Bulletin de Methodologie Sociologique*, 75, 24-39.
- Birch, S., Allen, N.J., & Sarmiento-Mirwaldt, K. (2017). Anger, Anxiety and Corruption Perceptions: Evidence from France. *Political Studies*, 65(4), 893-911.
- Delios, A., Malesky, E.J., Yu, S. & Riddler, G. (2024). Methodological errors in corruption research: Recommendations for future research. *Journal of International Business Studies*, 55, 235-251.
- Dickel, P. & Graeff, P. (2016). Applying Factorial Surveys for Analyzing Complex, Morally Challenging and Sensitive Topics in Entrepreneurship Research - The Case of Entrepreneurial Ethics. In E.S.C. Berger & A. Kuckertz (Hrsg.), *Complexity in entrepreneurship, Innovation and Technology Research - Applications of Emergent*

- and Neglected Methods FGF Studies in Small Business and Entrepreneurship. Springer, Berlin, 1-20.
- Dickel, P. & Graeff, P. (2018). Entrepreneurs' propensity for corruption: A vignette-based factorial survey. *Journal of Business Research*, 89, 77-86.
- Dunkelmann, F. (2022). Korruption und Corona. Mögliche Gemeinsamkeiten zweier Befragungsthemen. In S. Wolf & P. Graeff (Hrsg.), *Corona und Korruption. Gesellschaftswissenschaftliche Analysen*, Springer VS, doi: 10.1007/978-3-658-35664-4_7.
- De Waele, L., Weißmüller, K.S., & Van Witteloostuijn, A. (2021). Bribery and the Role of Public Service Motivation and Social Value Orientation: A Multi-Site Experimental Study in Belgium, Germany and the Netherlands, *Frontiers in Psychology*, 12:655964, doi: 10.3389/fpsyg.2021.655964.
- Dungan, J. A., Young, L. & Waytz, A. (2019). The Power of Moral Concerns in Predicting Whistleblowing Decisions. *Journal of Experimental Social Psychology* 85: 1–12.
- Fisman, R. & Golden, M.A. (2017). Corruption: What everyone needs to know. New York: Oxford University Press.
- Graeff, P. (2016). Social Mechanisms of Corruption: Analytical Sociology and its applicability to corruption research. *Analyse/Kritik*, 37(1), 53-71.
- Graeff, P. (2019). Begriffsverwendungen und Korruptionsverständnisse in der Forschung: Ein Vergleich von Definitionen und Ansätzen. In: P. Graeff & T. Rabl (Hrsg.): *Was ist Korruption? Begriffe, Grundlagen und Perspektiven gesellschaftswissenschaftlicher Korruptionsforschung*. 2. Auflage. Nomos, Baden-Baden, 271-298.
- Graeff, P., Sattler, S., Sauer, C. & Mehlkop, G. (2014). Incentives and prohibitions of abusing academic positions. Determining students tendencies to consent to bribing offers: a vignette study. *European Sociological Review*, 30(2), 230–241.
- Jensen, N. M., Li, Q. & Rahman, A. (2010). Understanding corruption and firm responses in cross-national firm-level surveys. *Journal of International Business Studies*, 41, 1481-1504.

- Kleinewiese, J. (2022). Good people commit bad deeds together. A Factorial Survey on the moral antecedents of situational deviance in peer groups. *Deviant Behavior*, 43(11), 1420–1431.
- Kleinewiese, J. & Graeff, P. (2020). Ethical decisions between conflicting priorities of legality and group loyalty: Scrutinizing the „Code of Silence“ among volunteer firefighters with a Vignette-based Factorial Survey. *Deviant Behavior*, 42(10), 1228-1241.
- Krumpal, I. & Voss, T. (2020). Sensitive questions and trust: Explaining respondents' behavior in randomized response surveys. *Sage Open*, 10(3), doi: 10.1177/2158244020936223.
- Leitgöb, H. & Wolbring, T. (2021). Die Methoden der sozialwissenschaftlichen Datenerhebung im digitalen Zeitalter. Entwicklungen, Möglichkeiten, Herausforderungen, In T. Wolbring, H. Leitgöb, & F. Faulbaum (Hrsg.) *Sozialwissenschaftliche Datenerhebung im digitalen Zeitalter. Schriftenreihe der ASI - Arbeitsgemeinschaft Sozialwissenschaftlicher Institute*. Springer VS, Wiesbaden, doi: 10.1007/978-3-658-34396-5_1.
- Lensvelt-Mulders, G. J. L. M. & Boeije, H. R. (2007). Evaluating compliance with a computer assisted randomized response technique: a qualitative study into the origins of lying and cheating. *Computers in Human Behavior* 23, 591-608.
- Liebe, U., Moumouni, I.M. Bigler, C. Ingabire, C. & Bieri, S. (2020). Using factorial survey experiments to measure attitudes, social norms, and fairness concerns in developing countries. *Sociological Methods & Research*, 49(1), 161-192.
- Mutz, D. C. (2011). *Population-Based Survey Experiments*, Princeton University Press, Princeton, doi: 10.1515/9781400840489.
- Niehaus, H. (2019). Annäherungen an einen Korruptionsbegriff des (deutschen) Strafrechts. In: P. Graeff & T. Rabl (Hrsg.): *Was ist Korruption? Begriffe, Grundlagen und Perspektiven gesellschaftswissenschaftlicher Korruptionsforschung*. 2. Auflage. Nomos, Baden-Baden, 221-234.
- Oliveros, V. & Gingerich, D.W. (2020). Lying about Corruption in Surveys: evidence from a Joint Response Model. *International Journal of Public Opinion Research*, 32(2), 384-395.

- Persson, A., Rothstein, B. & Teorell, J. (2013). Why anticorruption reforms fail – systematic corruption as a collective action problem. *Governance: An International Journal of Policy, Administration, and Institutions*, 26(3), 449-471.
- Rabl, T. (2019). Korruption aus der Perspektive der Psychologie. In: P. Graeff & T. Rabl (Hrsg.): *Was ist Korruption? Begriffe, Grundlagen und Perspektiven gesellschaftswissenschaftlicher Korruptionsforschung*. 2. Auflage. Nomos, Baden-Baden, 77-102.
- Schweitzer, H. (2022). Korruption. Eine sozialwissenschaftliche Darstellung und Analyse. Hamburg: Disserta Verlag.
- Tourangeau, R. & Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin*, 133(5), 859–883.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Vera, S.B. (2020). Accepting or resisting? Citizen responses to corruption across varying levels of competence and corruption prevalence. *Political Studies*, 68(3), 653-670.
- Wallander, L. (2009). 25 years of factorial surveys in sociology: A review. *Social Science Research*, 38, 505–520.

4 THIRD PAPER

Disentangling Differences in Perceived Sensitivity and Misreporting Behaviour: A Factorial Survey

Introduction

Sociological research frequently gathers information through surveys on private behaviours or attitudes, including those that are illegal, unsocial or sensitive. Examples of such behaviours include drug use (Biemer & Brown, 2005; Druckman et al., 2015), racism (Kuklinski et al., 1997), xenophobia (Krumpal, 2012), sexual activities (Roe-Sepowitz et al., 2019), and voting (Holbrook & Krosnick, 2010). For empirical research in many fields of social science the quality of data is crucial. In most cases, researchers are forced to rely on self-reports of respondents because objective information is not available (Gnambs & Kaspar, 2015, p. 1238). However, accumulated evidence in survey methodology suggests that responses regarding sensitive issues often fail to reflect what is called the 'truth' (Krumpal, 2013; Tourangeau & Yan, 2007). For instance, participants may provide inaccurate reports of behaviours that contradict widely accepted social norms or legal regulations due to concerns about potential negative consequences or embarrassment. Respondents tend to provide lower estimates for socially unaccepted behaviours and higher estimates for desirable behaviours or attitudes (e.g., Trappmann et al., 2014, p. 58f).

Misreporting or systematic nonresponse might cause inaccurate survey estimates (Tourangeau et al., 2000, p. 264f; Wolter & Preisendörfer, 2013, p. 322; Celhay et al., 2022, p. 1). Although many studies have investigated the sensitivity of survey topics and the accuracy of estimates, it is still relevant to consider Tourangeau's summary (2018, p. 506): *“On the causes side, it is not very clear when people are inclined to misreport in surveys. Kreuter, Presser, and Tourangeau (2008) found that the level of sensitivity of the question related to the level of misreporting. It would be useful to understand what questions people find to be sensitive.”*

This article aims to explore the sensitivity for additional topics as well as the relationship between respondents' sensitivity assessments and their potential response behaviour (nonresponse as well as intentional misreporting). To this end, a methodological approach is used that has not been applied to these aspects in the past.

Previous studies have investigated the extent to which misreporting can be assumed for various topics due to sensitivity. These studies have focused on topics that were categorised as 'sensitive' by researchers. This article uses a different approach. German respondents completed a factorial survey to evaluate their individual level of sensitivity in response to a hypothetical scenario. The scenario described a survey situation with varying aspects. The design simultaneously estimates factors that have been considered separately in the sensitivity literature. In this way, the multidimensional phenomenon of sensitivity assessments is taken into account (Mutz, 2011, p. 54).

Furthermore, sensitivity of the question and sensitivity of the answer are distinguished in the literature (Tourangeau, 2018, p. 503). This distinction leads to an understanding of sensitivity which could also differ between respondents. The design of this study implements this theoretical idea. Respondents were asked directly for an individual assessment, as the sensitivity assessments may differ between the researcher and the respondents.

The survey also included questions about respondents' assessment of their own response behaviour, specifically the probability of intentional misreporting and nonresponse. In this way, the probability of nonresponse and intentional misreporting can be related to the intensity of sensitivity assessment. This study design provides insight into the relationship between sensitivity assessments and response behaviour without directly exposing respondents to a potentially sensitive situation. This survey method has the advantage of reducing social desirability and sensitivity in the data. Additionally, respondents assessed the sensitivity themselves rather than researchers determining which topics are sensitive.

The overall structure of the article takes the form of six chapters. The following section outlines the theoretical foundations and derives hypotheses (chapter 2). The third chapter provides a short overview of the methodology of factorial surveys, the questionnaire, and the data. Chapter four presents the data analysis. Chapter five summarizes the results. The final chapter deals with the limitations of the approach and makes suggestions for future research (chapter 6).

Theory and hypotheses

The theoretical considerations are divided into two sections. First, the factors that influence the assessment of sensitivity from the respondents' perspective are identified from the literature. Subsequently, factors are derived that influence respondents' response behaviour, specifically intentional misreporting.

Response process and sensitivity

When respondents answer survey items, the accuracy of the resulting data depends on their ability and willingness to perform the necessary response process effectively (Tourangeau et al., 2000, p. 2; Krosnick, 1991, p. 214). Upon reviewing the research literature, the response process is ideally described as a step-by-step procedure that respondents undergo more or less straightforward. More precisely, Tourangeau, Rips, and Rasinski (2000, p. 7-16) detailed the process of answering a question through four key components: 1. Respondents need to carefully interpret the meaning of each question, 2. Respondents have to retrieve all relevant information from their memories, 3. Respondents should thoughtfully integrate that information into concise assessments, and 4. Respondents articulate those assessments in a manner that conveys meaning with clarity and precision. Proceeding carefully through every stage is commonly known as optimizing (Krosnick, 1991, p.214; Vannette & Krosnick, 2014, p. 313).

Although researchers expect that survey respondents optimize their response process for each question, the actual process often fails to meet these ideal expectations (Vannette & Krosnick, 2014, p. 313). Depending on the survey question (e.g., item content or wording) and respondent characteristics (e.g., motivation and clarity of opinion), certain steps may be deemed unnecessary or respondents may skip some components for other reasons (Tourangeau et al., 2000, p. 15; Krosnick, 1991, p. 215). According to Krosnick (1991, p. 215), a shift in response strategy from optimizing to so-called satisficing occurs if participants lose motivation to exert cognitive effort during a survey. Satisficing varies from only simplifying some steps of the response process (weak satisficing) to skipping one or more of the four major components (strong satisficing) (Krosnick, 1991, p. 214-220; Krosnick, 1999, p. 547f). Omitting or skipping cognitive processes can have a negative impact on data quality if, for

instance, respondents do not understand the question, cannot recall relevant information or are unable to formulate an appropriate response (Krosnick, 1991, p. 214; Vriesema & Gehlbach, 2021, p. 619; Celhay et al., 2022, p. 26).

In addition to these potential response biases, there may be other reasons why self-reports do not accurately reflect actual behaviours or attitudes. This article concentrates specifically on the effect of sensitivity on the accuracy of self-reported data. The most widely accepted and standard definition of sensitivity originates from Tourangeau and Yan (2007). They identify three dimensions of sensitivity: intrusiveness, threat of disclosure, and social desirability. These three dimensions refer either to the sensitivity of a *question* or the sensitivity of an *answer* (Tourangeau, 2018, p. 503).

According to Tourangeau and Yan's (2007) first dimension of sensitivity, participants may perceive survey *questions* as invasive or inherently offensive if they delve into social taboos (Tourangeau, 2018, p. 503). Such sensitive subjects are usually avoided in regular conversations and respondents may find them inappropriate to ask in a survey interview or questionnaire: *"[These questions] are seen as an invasion of privacy, regardless of what the correct answer for the respondent is. [...] Questions in this category risk offending all respondents, regardless of their status on the variable in question"* (Tourangeau & Yan, 2007, p. 860). Tourangeau (2018, p. 501) adds (based on (Singer et al., 1993) that in these cases *"privacy refers to unwillingness to reveal information at all. Respondents who say (or feel) that 'It's none of your business', are exhibiting privacy concerns."* Due to this circumstance, it is expected that such questions may result in higher nonresponse rates or drop-out rates, regardless of the respondents' actual answers. With regard to the factors influencing the respondent's sensitivity assessments, one can develop the first hypothesis (also see Blair et al., 2020; Tourangeau & Yan, 2007; Lee & Renzetti, 1990; Krumpal, 2013):

H1: The sensitivity assessment of respondents in a survey situation is substantially influenced by the survey item.

The second dimension by Tourangeau and Yan is known as the threat of disclosure (also see Lee & Renzetti, 1990, p. 511f). This threat relates to *answers* that could lead to negative consequences for respondents if their answer becomes known to external parties, which may include individuals or organisations (Tourangeau, 2018, p. 503). Depending on their response to a question, respondents may be reluctant to provide an honest answer. The level of

sensitivity is related on the response options and is influenced by whether respondents have anything to hide (Tourangeau & Yan, 2007, p. 860). For instance, a survey question addressing corrupt practices creates a threat of disclosure for individuals involved in corruption because they have something to hide. On the contrary, those who are not involved in corruption have nothing to hide, so the risk of exposure is non-existent for them. As demonstrated by this brief example, respondents could give a truthful response, but this decision might depend on their actual answer. Respondents with something to hide, could also decide not to tell the truth (Delios et al., 2024, p. 238). This facet of sensitivity has the potential to introduce bias into survey results (Cobo et al., 2021, p. 10; Preisendörfer & Wolter, 2014, p. 127; Krumpal, 2013, p. 2028 ff).

“Apart from questions that are inherently offensive and those that raise concerns about disclosure to third parties, there is a third type of sensitive questions – those that raise social desirability concerns. With these questions, there is a socially approved answer and a socially disapproved answer” (Tourangeau, 2018, p. 503). This third dimension of sensitivity by Tourangeau and Yan (2007) is also related to specific *responses* to a survey question. This dimension of sensitivity assumes the existence of clear social norms for a given behaviour or attitude. Responses that conform to these norms are considered socially desirable, while those that deviate from them are seen as socially undesirable (Tourangeau & Yan, 2007, p. 860; also see DeMaio, 1984, p. 258; Holtgraves, 2004, p. 161). Compliant participants do not have an incentive to modify their responses. The sensitivity once again relies on the true status of participants. This dimension of sensitivity can also introduce bias if respondents with a potentially undesirable behaviour or attitude refrain from admitting a violation of the social norm (also see Coutts & Jann, 2011, p. 170; Trappmann et al., 2014, p. 59; Krumpal, 2013, p. 2028; DeMaio, 1984, p. 257).

The theoretical consideration of the second and third dimension emphasise a difficulty in collecting sensitive information: The sensitivity assessment of respondents could depend on their true answer (Lee & Renzetti, 1990, p. 513). In most cases, however, the true answer cannot be known (Krosnick, 2018, p. 347; Gnambs & Kaspar, 2015, p. 1238).³¹ For topics where the answers are likely to be sensitive, researchers try to find ways of dealing with this.

³¹ In social science research, validation information is rarely available. In some cases, validation studies were possible, for instance when researching the nonresponse and misreporting behaviour of graduates of a university (Kreuter et al., 2008), the misreporting of voters for a referendum (Rosenfeld et al., 2016), misreporting in the

Some approaches analyse shifts in the survey setting, as the example of corruption research shows: *“However, it is possible that response bias on sensitive topics such as corruption may be lower when the survey is not conducted face-to-face”* (Richards, 2017, p. 5). Research shows that the method of survey administration is important when researching sensitive issues (Krumpal, 2013, p. 2034ff; Tourangeau et al., 2000, p. 295f; Tourangeau & Yan, 2007, p. 863).

Especially the existence of an interviewer (in face-to-face or telephone interviews) might create a considerable social reference in contrast to online surveys or postal surveys (Blair et al., 2020, p. 1299; Tourangeau, 2018, p. 138). Respondents may attempt to present themselves positively to interviewers and/or avoid feeling ashamed in front of an interviewer by giving a socially undesirable response or admitting illegal behaviour even if there is no personal relationship with the interviewer (Simpser, 2017, p. 8).

These studies look at the response behaviour of respondents and compare this between different survey modes (e.g., Kreuter et al., 2008). Differences in response behaviour between the conditions are then attributed to the influence of sensitivity or social desirability (Gnambs & Kaspar, 2015, p. 1254). The design of this study is intended to test whether the presence of an interviewer already increases the sensitivity assessments. A theoretical justification for a change in sensitivity by the presence of an interviewer is based on dimensions two (treat of disclosure) and three (social desirability) by Tourangeau and Yan (2007). As outlined, Tourangeau and Yan assume that sensitivity may arise from respondents’ fear that their answer could become known to others (Tourangeau & Yan, 2007, p. 860). The presence of an interviewer during a survey could increase this fear of disclosure and thus also lead to an increased sensitivity assessment of respondents. The presence of an interviewer may also emphasise social desirability. The sensitivity assessment of respondents could therefore be increased by the presence of an interviewer:

H2: The sensitivity assessment increases when the survey situation involves an > interviewer (in person or on the telephone).

Tourangeau (2018, p. 501) also states that interviewees may have another concern: *“In the survey setting, respondents may worry that someone else in the household may overhear what*

area of delinquent behaviour in a sample of criminally convicted persons (Wolter & Preisendörfer, 2013; Preisendörfer & Wolter, 2014) or misreporting of abortion compared to validation from an abortion clinic (Rasinski et al., 1999).

they report in an interview". In particular, the presence of third parties during an interview is not uncommon. In their meta-analysis Tourangeau and Yan (2007, p. 868) also examined the influence of third-party presence. They found a large effect of parental presence when young people are interviewed.

In another study, Mneimneh et al. (2015) analysed data from a survey about mental health. The authors compared the response behaviour to items described as sensitive or neutral depending on the presence of third parties (no differentiation between different bystanders, data was collected through interviews). The authors found that the reporting of sensitive information is influenced (among other things) by the presence of a third party (Mneimneh et al., 2015, p. 693). The studies focus on the presence of bystanders or third parties on the respondent's response behaviour. As with regard to a possible interviewer effect, the presence of third parties will now be analysed in terms of sensitivity assessments. The relevance of the presence of bystanders can be derived from the empirical studies (also see Diop et al., 2015).

Furthermore, a theoretical justification for the change in sensitivity caused by bystanders is also possible if one considers the 'threat of disclosure' and 'social desirability' dimensions of sensitivity. As already outlined, Tourangeau and Yan (2007, p. 860) assume that respondents' sensitivity assessments may arise from the fear that their answer will become known to others. The presence of bystanders during a survey could increase this fear of disclosure. The presence of bystanders could also increase the social desirability. Thus, the presence of a third party leads to an increased sensitivity assessment of respondents. The empirical part of this paper analyses two specific examples of the presence of third parties, which are assumed to be realistic for the majority of respondents. Family members or partners within earshot may be a relevant social referent whose presence might have an impact on the assessment of sensitivity (Blair et al., 2020, p. 1299):

H3: The sensitivity assessment increases when family members or partners are present in the survey situation.

Sensitivity, anonymity and misreporting in surveys

In summary, the sensitivity of a survey question or situation can result in different actions: (1) respondents may refuse to answer a survey question (item nonresponse), (2) respondents may edit their answer during the reporting process (intentional misreporting), or (3) respondents may answer truthfully despite the sensitivity of the question or situation.³²

There are two ways of describing the issue of item nonresponse: On the one hand, respondents may choose to avoid answering sensitive questions (Krumpal, 2013, p. 2026; Couper et al., 2008; Tourangeau, 2018, p. 504). Researchers can observe this nonresponse rate and some use the nonresponse rate as a criterion for the sensitivity of a survey item (Yan, 2021, p. 114; Tourangeau & Yan, 2007, p. 861). The nonresponse rate is identified as a measure of bias in available data. On the other hand, a refusal to answer can be uncomfortable because it could be understood as a kind of communicative signal from the interviewee. Respondents may think that refusing to answer questions during an interview that pertains to norm violations or breaking the law can be perceived as an admission of guilt (Wolter, 2019, p. 171). This is especially true in sensitive situations if respondents are driven by social desirability or fear of disclosure. In such cases, refusing to answer may be less attractive or perceived impossible as a response strategy.

In addition to (visible) refusal to answer, respondents can also choose to misreport. This article focuses on the phenomenon of deliberate or intentional misreporting³³ opposed to unconscious response biases or automated response behaviour (Tourangeau et al., 2000, p. 14). *“Misreports are conceptualized as a motivated process under which respondents edit their answers before they report them”* (Böckenholt, 2014, p. 515). Following the idea of a step by step response process, deliberate misreporting is not connected to failure or lack of knowledge by respondents (e.g., memory gaps). Rather, these respondents understand the item (step 1), retrieve relevant information (step 2), use this information to make a judgement on the question (step 3) and then consciously decide not to give the answer generated in the previous steps. To make it easier to distinguish, the possible answer can be labelled as a ‘preliminary answer’, as long as the respondent has not yet given it. The answer given may

³² This study does not focus on unit nonresponse and possible resulting problems for the data quality. For interested readers, the publications by Yan (2021) can be a good introduction.

³³ Deliberate misreporting and intentional misreporting are used synonymously.

then either match the preliminary answer or, in the case of intentional misreporting, deliberately deviate from it. Intentional misreporting thus characterises respondents' response behaviour (at the end of the response process) and is not a term for the process itself.

As already described in the previous chapter, differences in the response behaviour are often attributed to the influence of sensitivity or social desirability (e.g., Gnambs & Kaspar, 2015, p. 1254; Kreuter et al., 2008). Tourangeau and Yan (2007) already address the response behaviour of respondents in their theoretical differentiation of the dimensions of sensitivity. The authors assume that the 'threat of disclosure' and social desirability have the potential to result in misreporting (Tourangeau & Yan, 2007, p. 863). Respondents who have something to hide are likely to do so. Respondents whose preliminary answer is socially undesirable would also have an incentive to hide it. In this way, it is expected that a higher sensitivity assessment of respondents leads to a higher probability of intentional misreporting. Empirical studies on misreporting also establish this relationship and assume that sensitivity might bias resulting data (Krumpal, 2013; Tourangeau & Yan, 2007, p. 863; Kuhn & Vivyan, 2021, p. 384; Gnambs & Kaspar, 2015, p. 1248):

H4: An increased sensitivity assessment increases the probability of intentional misreporting.

Researchers dealing with sensitivity want to make misreporting due to sensitivity less likely. A fundamental idea is linked to the 'threat of disclosure' and 'social desirability' dimensions of Tourangeau and Yan (2007). Tourangeau (2018, p. 501) writes particularly with regard to the threat of disclosure: *"In the survey setting, respondents may worry that [...] another federal agency or a corporation would learn what they said."* As a logical consequence, researchers are trying to reduce or completely eliminate this concern.

Research has strongly been focused on ensuring anonymity (Krumpal, 2013, p. 2041; Vannette, 2018, p. 336). Anonymity means that the respondents cannot be identified after the survey. Accordingly, it should not be possible to trace individual responses back to respondents (Lelkes et al., 2012). A classic approach is to ensure anonymity at the beginning of a questionnaire or interview (Vannette, 2018, p. 336). Another way is to use special, indirect questioning techniques. Well-known examples are the Randomised Response Technique

(RRT) or the Item Count Technique (ICT). The main objective of indirect forms of questioning is to improve the anonymity.

This can be illustrated with the Item Count Technique: This survey method incorporates a sensitive item within a list of items. This list of items is answered with one collective response (usually with the sum of the agreed statements) (Glynn, 2013). The purpose is to enhance the anonymity of the response, as interviewers or researchers analysing the response cannot directly identify the individual's response to the sensitive item as long as the respondent does not agree with all or none of the items (Hinsley et al., 2019; Droitcour et al., 1991; Glynn, 2013). Reasoning is also possible using the Randomized Response Technique: A random mechanism is used to decide whether a respondent answers a sensitive or an alternative item. The response is given without revealing the outcome of the randomisation mechanism. In this way, the respondents are able to match the answer to the question, whereas the researchers use statistical methods to estimate the answer to the sensitive item (Warner, 1965). The primary objective of these indirect questioning modes is to improve privacy protection in order to reduce the influence of all sensitivity dimensions (Krumpal & Voss, 2020, p. 5). This, in turn, aims to improve data quality (Buchanan, 2000, p. 123).

Taking this idea further, researchers do not assume that the sensitivity of a question or topic can be completely eliminated. Indeed, anonymity plays a moderating role that alters the influence of sensitivity on respondents' response behaviour.

These procedures as well as the anonymity assurances do not initially affect the sensitivity of the item of interest. However, the increased anonymity of the individual response within the survey situation shall reduce the probability of intentional misreporting. Despite the assessed sensitivity of the matter, misreporting is less likely when anonymity is provided. This leads to two hypotheses modelling a linear influence of anonymity as well as an interaction effect incorporating anonymity and sensitivity³⁴:

³⁴ As one anonymous reviewer noted, modelling the anonymity assessment as an influencing factor on the sensitivity assessment or as a moderating effect for the influence of the interviewer/bystander variable on the sensitivity assessment would also be possible. In the case of ICT or RRT, it is no longer possible for the interviewer or bystander to understand which specific answer is given due to the randomisation mechanism. In these cases, the questioning technique (understood as a significant increase in the anonymity of the answers) could therefore have a moderating effect on the sensitivity assessment. At this point, the focus is on whether the anonymity assessment of respondents can make misreporting less likely, even though respondents assess the content as sensitive in a traditional survey.

H5: An increased anonymity assessment decreases the probability of intentional misreporting.

H6: An increase in anonymity can reduce the influence of the sensitivity assessment on the probability of intentionally misreporting.

The assumption of an interaction effect is also tested in the analysis. In the past, ensuring anonymity in surveys has often been achieved through explicitly assuring anonymity or utilising specific survey methods. In the context of this study, the participants' assessments of anonymity will be taken into account. It is assumed that when a survey method is subjectively evaluated as more anonymous, the probability of misreporting will decrease and the effect of sensitivity can be moderated.

Methods, operationalisation and data

The third chapter of this article briefly introduces the method of factorial surveys (3.1), the survey design of this study (3.2), and the database (3.3) for the analysis that will follow in chapter four.

Factorial surveys

In order to test the hypotheses, a factorial survey was carried out. German respondents were confronted with descriptions of hypothetical situations which are called vignettes. Each vignette contains a mixture of several different attributes simultaneously, and these attributes (the so-called vignette dimensions) are varied experimentally. Factorial surveys are characterised by a combination of survey and experiment. The analysis of vignettes aims to achieve internal as well as external validity. The internal validity is based on the experimental variation of the vignette dimensions whereas the external validity arises from conducting this experiment in a population sample (Mutz, 2011, p. 5; Schlüter & Schmidt, 2010, p. 93; Treischl & Wolbring, 2022, p. 145).

After reading a vignette, respondents should evaluate it as if they were experiencing the scenario (Auspurg et al., 2014, p. 137). This indirect method is argued to be particularly

suitable for assessing personal opinions. Furthermore, Factorial Surveys are less likely affected by social desirability bias than standard interview techniques because respondents only respond as if they were in the situation and evaluate the questions of interest in an indirect way (Wallander, 2009, p. 506; Mutz, 2011, p.64).

Auspurg et al. (2014, p. 137) connect this with the multidimensional design: *“The multifactorial design in which several dimensions vary at the same time forces respondents to make trade-offs between single dimensions. Furthermore, as sensitive issues and attitudes are assessed indirectly, FSs [Factorial Surveys] are considered to be less susceptible to SDB [Social Desirability Bias] than direct questioning”*. The mentioned multifactorial design of vignettes makes it possible to analyse how the characteristics of the dimensions influence a dependent variable - in this article the sensitivity assessment (Atzmüller & Steiner, 2010, p. 129). Furthermore, this survey method can account for individual differences in respondents' sensitivity assessments. This corresponds to the theoretical considerations that the sensitivity is assessed differently by respondents depending on their preliminary answer (Tourangeau & Yan, 2007, p. 860; Tourangeau, 2018, p. 503).

Survey design

To test the hypotheses, the vignettes provide a description of a survey situation (see table 1). Respondents were asked whether they intended to respond to the survey question if they would be in the described situation. Secondly, respondents evaluated the sensitivity of the survey situation. Finally, participants were requested to assess their probability of responding truthfully to the survey question.

In order to reduce the potential for satisficing, research suggests to have a low level of task difficulty (Vannette & Krosnick, 2014, p. 320; Vannette, 2018, p. 333f). For this reason, respondents were not directly confronted with the concept of sensitivity. Instead, respondents were asked how uncomfortable they felt about the situation described in the vignette. In this way, sensitivity was expressed and measured in a simplified way.

The predicted misreporting behaviour was also not asked in a direct manner. It is assumed that a question on intentional lying or deliberate misreporting would lead to strong social desirability effects (despite the use of the factorial survey). Therefore, respondents were

asked *“How likely would you answer the question truthfully?”*. By reversing the probability of a truthful answer, this question operationalises the probability of intentional misreporting.

Table 1: Example of a vignette presented to respondents, including questions to assess

Please imagine the following situation:
You are taking part in an *online survey*.
No other people are present at the time.
You are asked the following question:
In the last 5 years, have you offered someone money to do something for you that no one else should know about?
☐ Yes
☐ No
☐ Prefer not to say

Would you answer this question?

- ☐ Yes
☐ No

How uncomfortable is the overall situation described above for you personally?

	Not uncomfortable at all										Very uncomfortable
	0	1	2	3	4	5	6	7	8	9	
I think this situation is ...	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

How likely would you answer the question truthfully?

	Very unlikely										Very likely
	0	1	2	3	4	5	6	7	8	9	
It is...	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

Note: This English version is translated from the original German questionnaire. The varying dimensions are shown in *italics*. The survey item in the vignette (in this example: the bribery question) was shown with answering option but respondents could not choose a response for this item because it was part of the vignette.

Three dimensions with different characteristics were included in the vignettes: survey item, survey mode, and presence of bystanders. Table 1 presents an exemplary vignette, and the appendix contains another table that provides an overview of the vignette dimensions, their levels, and the base category for later analysis (see table A1). The three vignette dimensions and their characteristics are explained in the following.

In terms of the 'survey item' dimension, a total of six items were selected, assuming varying degrees of sensitivity.³⁵ The selection of items was based on the dimensions of sensitivity according to Tourangeau and Yan (2007).

The item with the least expected sensitivity was the age of the respondent (in years). This item shows minimal nonresponse in population surveys, and no theoretical reasons for its sensitivity can be derived from the literature. The question cannot be categorised as a social taboo, nor are the answers conflictual.

The second item that may be sensitive for a small number of respondents in Germany was a question about religion. Cumulatively, more than 90% of Germans identify with Christianity *or* no religion. As these answers reflect societal norms, no sensitivity or intentional misreporting is expected for these individuals. However, for those who belong to less common religions or whose religion may be associated with prejudice in Germany (Kantar, 2019), the question about their religion may cause some sensitivity. Considering all respondents collectively, the sensitivity for this question was assumed to be slightly higher than that about respondents' age.

The third item requested about the respondents' sexual orientation. In general, questions about individuals' sexuality are often considered taboo. For instance, research on the number of sexual partners (Tourangeau et al., 2000, p. 297), the use of condoms (Haber et al., 2018, p. 4; Lépine et al., 2020, p. 409), or abortion (Huber-Krum et al., 2020, p. 4; Moseson et al., 2017, p. 7; Kissling & Jackson, 2022, p. 33) has shown biases and/or increased rates of nonresponse due to the sensitive nature of these topics (Smith, 2014, p. 1140-1144). Within this field, the issue of sexual orientation, as opposed to specific practices or habits, was assumed to be less sensitive for most people. The majority of Germans identify as heterosexual, and this has long been established in public perception as the (only) sexual norm. Homosexuality, bisexuality or other orientations deviate somewhat from this majority. Furthermore, the open expression of these sexual orientations can be associated with negative experiences (European Union Agency for Fundamental Rights, 2020). It is therefore

³⁵ Respondents were not asked to imagine that a certain answer option would apply to them. On the one hand, this practice was not considered to be very effective. Respondents might have to put themselves in an answer category that does not apply to them and this could result in a distorted response pattern rather than reliable data. On the other hand, this would be a *very* cognitively demanding task to put oneself in a survey situation with a specific item and a specific applicable answer. This complexity is unlikely to result in good data quality.

assumed that the question about sexual orientation may be associated with sensitivity for (some) non-heterosexual respondents.

In summary, the first three levels of the dimension 'survey item' are characterised by a low or selective degree of sensitivity. Conversely, the next three levels are expected to be more sensitive.

A question regarding experiences of violence (*“Have you been the victim of physical violence by anyone in the past 12 months?”*) is closely linked to feelings of shame and stigma, and may also be associated with admitting weakness. In general, the experience of violence may be uncomfortable for respondents. For instance, they may blame themselves for the experience or feel uncomfortable about not being able to prevent a situation of physical violence. For these respondents, social desirability could play a more important role and those who have experienced violence may prefer not to admit it. The sensitivity in this case therefore lies in a certain response, the admission of experienced violence. However, it cannot be assumed that dimension two, the threat of disclosure, plays a major role in this case, as this is a passive experience of violence rather than an active one. Sanctions are often linked to the exercise of violence whereas the experience of violence is not sanctioned. According to the definition by Tourangeau and Yan (2007), the dimension of social taboo is also not applicable in this case, as the question itself would have to trigger rejection. This cannot be assumed for respondents who have not experienced violence.

The next question about infidelity in the current relationship is subject to strong social norms supporting monogamy and fidelity in relationships. Even if some people have open relationships, admitting infidelity within a relationship would be against the clear social norm and would also carry negative sanctions (e.g., expected disapproval from interviewers or bystanders). Admitting infidelity could be categorised as socially undesirable and also associated with the threat of disclosure, based on Tourangeau and Yan's dimensions (2007). Additionally, some respondents may consider even asking about infidelity in a relationship to be impertinent, as it may violate deeply held or even religious principles. Finally, it should also be emphasised that infidelity is not a criminal offence in Germany.

In this respect, the sixth item might represent a further increase in sensitivity. This question relates to active bribing (*“In the last 5 years, have you offered someone money to do something for you that no one else should know about?”*). Given the respondents from Germany, bribery

represents an increase in sensitivity, as this behaviour is not only socially sanctioned. Bribery is a violation of the principle of fairness and is against the law. Bribery is considered a socially undesirable behaviour and sensitive according to the third dimension of Tourangeau and Yan's (2007) framework. Furthermore, bribery is explicitly illegal, prohibited and punishable by law in Germany. By definition, bribery must remain hidden (Delios et al., 2024, p. 238). Therefore, agreeing to this point is associated with a very high level of sensitivity due to the threat of disclosure as well.

The dimensions 'interview mode' and 'presence of bystanders' each have three levels representing different degrees of the presence of an unknown interviewer as a social reference and the presence of family members or partners as the attendance of known social references in the interview situation (Blair et al., 2020, p. 1299).

The combinations of all described vignette dimensions with the corresponding levels results in $6 \times 3 \times 3 = 54$ combinations. As this universe of vignettes could not be judged by every respondent, a random selection (without replacement) was used (Treischl & Wolbring, 2022, p. 154; Jasso, 2006, p. 343; Atzmüller & Steiner, 2010). Each respondent was given four vignettes to evaluate. The order of these four vignettes was completely random and all vignettes are drawn evenly distributed. The vignettes were included in the middle of the questionnaire.

Each questionnaire contained eight demographic questions, four vignettes and some additional questions. The following demographic information were collected: age, gender, being in a current relationship, sexual orientation, education, employment and religion. After these demographic questions, the four vignettes were presented. Following the vignettes, respondents were asked to rate the anonymity of responses in web surveys, telephone surveys and face-to-face surveys as well as their experience with these types of surveys. In addition, an instructional manipulation check (IMC) was included in the questionnaire. All these items can be found in the appendix (see table A2). IMCs aim to assess whether respondents read the instructions at all and/or as carefully as they should. Easy and short IMC questions shall detect severe cases of lack of attention. The failure to answer IMCs correctly was described as a strong form of satisficing (Anduiza & Galais, 2017). Response to the IMC item is included as an additional control variable in the analysis, given the particular

importance of respondents carefully reading the experimental design instructions and rating each vignette.

Data

The fieldwork for the factorial survey was carried out over a two-week period in July 2023. The data collection was carried out online by the Norstat Panel. A combined quota for age and education was used to select the respondents. The target German population was between 18 and 55 years old. After collecting the data, it was prepared for analysis. To show the probability of intentional misreporting, the truthfulness scale was inverted. As previously stated, this approach was chosen to reduce the difficulty of the task for respondents as well as the negative effects of social desirability. The anonymity assessment scale is transformed in the same way. Furthermore, the anonymity assessment of the respondents was used so that this referred to the mode described in the vignette. For instance, if the vignette described the setting of an online survey, the inverse of the respondent's anonymity assessment for online surveys was used as the data point for assessment of the violation of anonymity. The control variable for survey experience was also used depending on the mode of the vignette scenario, so that the respondent's experience with the described setting was covered. The demographic variable of education was edited. The original item allowed for very detailed information and was summarised in three groups (low education, medium education and high education). This three-point scale is based on the European Qualifications Framework (EQF) established by the European Union (Europass by European Union). The response to the IMC was recoded to a dummy variable, as only one of the four response options aligned with the instruction (and therefore represents a successfully passed IMC). The other three response options were summarised as "not successfully passed IMC".

Furthermore, the cases were standardised. 67 respondents who dropped out of the questionnaire and 5 respondents which started the questionnaire twice (one incomplete and one complete participation) were excluded from the analysis sample. Occasional item nonresponse did not lead to the exclusion of respondents from the dataset. The reduced sample size includes 1640 respondents. Of those completing the questionnaire, 52% were female and 47% male (less than 1% diverse). 76% of respondents were heterosexual and 67% were currently in a relationship. The average age was 40. The majority of respondents were

either non-religious (44%) or Christian (49%). The quota sampling aimed to achieve an equal distribution of educational levels based on school diplomas. The exclusion of some cases resulted in 31% of respondents with a low (EQF Level 1 and 2), 35% with a medium (EQF Level 3) and 34% with a high level of education (EQF Level 4).

In a factorial survey, the orthogonality (equal occurrence of each vignette dimension among each other) and the balance of the vignette dimensions can be considered before analysing the data. The vignette universe as a whole satisfies these two criteria by design. The random assignment of vignettes to participants and drop-outs could, at least in theory, affect orthogonality and balance. This is going to be briefly explored to prepare for analysing the data. Generally, using a larger number of vignette samples improves the orthogonality and balance of the vignette universe, increasing the probability of accurately estimating relevant main and interaction effects (Dülmer, 2016, p. 37). Considering the total of 54 vignettes in this case, each vignette was answered between 117 and 126 times (variations due to the random selection of vignettes and drop-outs). Furthermore, the bivariate correlations between the vignette dimensions can be analysed (see table A3 in the appendix). The correlations are both low and insignificant indicating orthogonality. Chi-square tests were also carried out between the vignette dimensions showing no significance (see table A3 in the appendix). This means that certain characteristics of one vignette dimension did not systematically occur more frequently with characteristics of another vignette dimension. The balance of the data was therefore not negatively affected by the drop-outs. Thus, the experimental factors can be regarded as independent of each other, which provides unbiased results for the following analysis.

Empirical results

The next section of this article will present the statistical analysis and is divided into two parts. The first part focuses on the sensitivity assessment as the dependent variable (section 'Sensitivity assessment'). Subsequently, the sensitivity is used – along with other independent variables – to explain the response behaviour of respondents (section 'Nonresponse and intentionally misreporting'). In particular, nonresponse and intentional misreporting are analysed.

Sensitivity assessment

To test the first three hypotheses regarding the factors influencing sensitivity assessments, four multivariate ANOVA tests were first conducted, as each respondent answered four vignettes. Table 2 presents the results of the ANOVA, indicating that the mean values of sensitivity assessments were significantly dependent on the survey item of the presented scenario. In contrast, the situational dimensions (presence of bystanders and survey mode) do not result in significant mean differences of the sensitivity assessments.

Table 2: Descriptive statistics and multivariate ANOVA

Vignette dimension		Mean	F-value	Vignette dimension		Mean	F-Value
First vignette per person				Second vignette per person			
Survey item	Age	1.55	29.72***	Survey item	Age	1.26	19.51***
	Religion	1.37			Religion	1.22	
	Sexual orientation	2.05			Sexual orientation	2.07	
	Infidelity	3.10			Infidelity	3.06	
	Experience of violence	3.68			Experience of violence	2.68	
	Bribery	2.09			Bribery	2.43	
Survey mode	Web Survey	2.20	0.77	Survey mode	Web Survey	2.10	0.31
	Telephone survey	2.41			Telephone survey	2.22	
	Personal interview	2.32			Personal interview	2.07	
Presence of bystanders	None	2.39	0.44	Presence of bystanders	None	2.10	0.12
	Partner	2.28			Partner	2.14	
	Family member	2.26			Family member	2.15	

Third vignette per person				Fourth vignette per person			
Vignette dimension		Mean	F-value	Vignette dimension		Mean	F-Value
Survey item	Age	1.13	15.63***	Survey item	Age	1.21	25.18***
	Religion	1.21			Religion	1.03	
	Sexual orientation	1.92			Sexual orientation	1.77	
	Infidelity	2.74			Infidelity	3.27	
	Experience of violence	2.66			Experience of violence	2.54	
	Bribery	2.30			Bribery	2.07	
Survey mode	Web Survey	2.04	0.21	Survey mode	Web Survey	1.86	0.67
	Telephone survey	1.96			Telephone survey	1.92	
	Personal interview	2.04			Personal interview	2.03	
Presence of bystanders	None	1.93	1.78	Presence of bystanders	None	1.87	2.52
	Partner	1.88			Partner	2.12	
	Family member	2.19			Family member	1.81	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This initial assessment, which is based on mean differences, is further analysed through a multivariate, multilevel regression model (see table 3). This is consistent with the hierarchical data structure (four vignettes per respondent) and illustrates different characteristics within the dimensions (Hox, 2010, p. 11; Wallander, 2009, p. 516).

Table 3: Multilevel analysis of sensitivity assessments

	(1)	(2)
Survey item	(ref.: Age) Religion 0.03 Sexual orientation 0.77*** Infidelity 1.82*** Experience of violence 1.55*** Bribery 1.02***	(ref.: Bribery) Age -1.02*** Religion -0.99*** Sexual orientation -0.25** Infidelity 0.80*** Experience of violence 0.53***
Survey mode	(ref.: Web Survey) Telephone survey 0.12 Personal interview 0.10	(ref.: Web Survey) Telephone survey 0.12 Personal interview 0.10
Presence of bystanders	(ref.: None) Partner 0.01 Family member 0.08	(ref.: None) Partner 0.01 Family member 0.08
Constant	1.13***	2.15***
Sd (Cons)	2.00	2.00
Sd (Residual)	1.97	1.97
N (Level 2, Respondents)	1 640	1 640
N (Level 1, Vignettes)	6 560	6 560

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In table 3, a positive sign of the coefficients indicates higher values of sensitivity. The presented models do not differ in terms of the independent variables used, namely the survey item, presence of bystanders, and survey mode. However, these variables were modelled differently. Model 1 estimates the influence of the vignette dimensions on the sensitivity assessment with age as the reference category for the survey item. In model 2 the reference category for the survey item is shifted from age to bribery.

The survey item described in the vignette (H1) has a significant influence on respondents' sensitivity assessments. Based on the two models, an order of sensitivity assessments can be identified that partially corresponds to the theoretically derived expectation. The questions regarding age and religion were found to be of low sensitivity. The two items do not differ significantly from each other (model 1), but show a clear difference to other items (model 2).

The question about sexual orientation is evaluated as significantly more sensitive, but shows a medium sensitivity compared to the other items. This could be due to the idea that respondents who do not belong to the heterosexual majority assess a higher sensitivity. A separate calculation shows a significant interaction between the respondents' own non-heterosexual orientation and the sensitivity rating for this item (see model 1 of table A4 in the

appendix). Non-heterosexual respondents (24% of the sample) rate the sensitivity of this question significantly higher. Interestingly, the conditional main effect of the sexual orientation item remains significant and positive in this model. This positive coefficient is estimated based on the modelling of the interaction term in such a way that it indicates the effect of the question on sexual orientation (only) for all heterosexual respondents. The positive sign and the remaining significance of the coefficient of the survey item of sexual orientation indicate that it is possibly a social taboo (as described by Tourangeau & Yan, 2007, p. 860) because the significant influence on the sensitivity assessment does not depend on the actual sexual orientation.³⁶

Continuing with the analysis of the first model in table 3, there is a deviation from the previously formulated assumption of medium sensitivity for the item on infidelity in a relationship. Infidelity in a partnership emerges as a highly sensitive issue in this survey, even surpassing the illegal conduct of bribery. It appears that the norm of faithfulness in a partnership, and the social desirability it engenders, have a higher impact on sensitivity evaluations than the possible violations of laws.

The bribery item, by contrast, turns out to be the item with the third highest sensitivity rating and is therefore defined as a reference category in model 2. This model highlights that the question about bribery can empirically be described as of medium sensitivity.

The remaining item, which asks about the respondent's experience of violence, can be categorised on the sensitivity assessment scale between bribery and infidelity. The experience of violence is rated significantly more sensitive than involvement in bribery. To summarise, it can be clearly seen that the survey item within the vignette scenarios led to significant differences with regard to the sensitivity assessment. H1 cannot be rejected.

The survey mode with the characteristics online survey, telephone survey and personal interview is the second independent variable of the models. According to H2, an interviewer effect is assumed. First, the three stages are distinguished. A telephone interview could possibly represent an intermediate stage, since the interviewer does not interact face-to-face

³⁶ It is important to acknowledge that biases related to sensitivity may also impact the direct question regarding the respondents' sexual orientation (which was asked at the beginning of the questionnaire). It is therefore conceivable that respondents who identified as heterosexual in the survey may have non-heterosexual orientations and may have concealed this.

with the respondent. However, table 3 shows no significant influence on the respondents' assessments of sensitivity across these three levels. To further test the hypothesis of the relevance of an interviewer effect, the mode variable was dichotomized in another calculation (see model 2 in table A4 in the appendix). This coefficient remains insignificant.

An interaction between the survey mode and the dimensions of the survey items is also considered to be conceivable. This interaction is calculated with both the three-level variable (survey mode) and the dichotomous interviewer variable (see models 3 and 4 in table A4 in the appendix). In model 3 (table A4), the dichotomous interviewer variable shows a significant interaction term with the bribery item. Surprisingly, the presence of an interviewer in the scenario reduces the respondent's sensitivity assessments. The model with the more differentiated distinction between online survey, telephone survey, and personal interview (see model 4 in table A4) indicates that the effect is exclusively attributable to the characteristic of the personal interview in combination with the bribery item. The coefficient for the bribery item in combination with the telephone survey is insignificant. The other items (age, religion, sexual orientation, experience of violence and infidelity) show neither significant interaction effects with the interviewer variable (model 3 in table A4) nor with the more differentiated distinction between online survey, telephone survey and personal interview (model 4 in table A4). Thus, in all models, the presence of an interviewer in the scenario has only a minor effect on the assessment of sensitivity. H2 has to be rejected for most survey topics.

The third independent variable is the presence of bystanders (H3). The mention of the presence of family members or partners shows no significant influence on the respondents' sensitivity assessments in either model 1 or model 2 of table 3. To investigate the possibility of bystanders having a reinforcing influence on some values of the survey item dimension (especially for the question about infidelity) an interaction term between the survey item and the bystander variable is also included in two additional models (see models 5 and 6 in table A4 in the appendix). A significant difference from the reference category (survey item: age or bribery; bystander: none) is observed for the survey item of infidelity. Both the presence of a family member and the presence of a partner in the interview situation increase sensitivity assessments beyond the conditional main effect. In contrast, the other combinations of characteristics show no significant change in respondents' sensitivity assessments.

Furthermore, the conditional main effect of the presence of a bystander remains insignificant despite the two modelling approaches. In summary, barely of the calculated models show an effect of the presence of family members or partners on the sensitivity assessments. Overall, H3 has to be rejected with these data.

In summary, the most models do not show significant differences with respect to the survey mode and the presence of bystanders. This analysis finds that questionnaire's contents seem to determine sensitivity which serves as a predictor of response behaviour in the next step of the analysis.

Nonresponse and intentionally misreporting

Before analysing the respondents' assessment of intentional misreporting, it is important to first address the question of nonresponse posed earlier in the questionnaire. This will provide context for the subsequent analysis and ensure a logical flow of information. Participants were requested to evaluate whether they would respond to the corresponding item in the presented survey situation (see table 1 for an illustration). To analyse this question, three multilevel logit models³⁷ are estimated. Table 4 includes models with the same or similar independent variables that are modelled differently.

³⁷ The dependent variable of these models is the dichotomous distinction between predicted refusal to answer and to give an answer. Across all vignettes, respondents indicated in 8.5% of cases that they would not want to give an answer.

Table 4: Multilevel analysis of nonresponse prediction

	(1)	(2)	(3)
Sensitivity assessment	0.54***		
Violation of Anonymity (ref: V. very unlikely)			
V. rather unlikely	0.39	0.37	0.37
V. rather likely	0.85**	0.94***	0.94***
V. very likely	0.94**	1.20**	1.20***
Education (ref: low)			
Medium	-0.06	-0.19	-0.19
High	-0.53*	-0.52*	-0.52*
Gender (ref.: male)			
Female	0.22	0.30	-0.19
Diverse	-0.05	0.53	-0.52*
Age (in years)	-0.01	-0.01	-0.19
Item (vignette dimension)		(ref.: age)	(ref.: bribery)
Age			-3.00***
Religion		1.05**	-1.95***
Sexual orientation		1.69***	-1.31***
Infidelity		2.72***	-0.28
Experience of violence		2.65***	-0.35
Bribery		3.00***	
Constant	-5.96***	-6.52***	-3.53***
N (Level 2, Respondents)	1 615	1 615	1 615
N (Level 1, Vignettes)	6 460	6 460	6 460

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Model 1 considers the influence of respondents' sensitivity assessment and their evaluation of the potential violation of the anonymity. In addition, the demographic variables education, gender and age were controlled for. This model shows a significant influence of respondents' sensitivity assessment on their prediction of nonresponse. As sensitivity increases, respondents are more likely to consider refusing to answer.

There is also a significant influence of the violation of anonymity. The difference between the first two levels of this scale (violation very unlikely to violation rather unlikely) is not statistically significant. However, when respondents rate the violation as rather likely or even very likely, they are significantly more likely to consider nonresponse as possible. The demographic variables show no significant differences with the exception of high education. Respondents with a high level of education are less likely to consider that they would refuse to answer than respondents with a low level of education.

In model 2, the sensitivity assessment variable is replaced by the specific items of the vignette. The other independent variables are the same as in model 1. The independent variables

included in model 1 continue to have significant effects. It is interesting to note that the coefficients of the higher degrees of violation of anonymity are more pronounced than in model 1. This reinforces the impression that respondents' assessments of the anonymity of their responses influence their response behaviour. Considering the survey items described in the vignette, it can be seen that each characteristic significantly increases the probability that respondents estimate nonresponse (relative to the reference category).

In model 3 the bribery item is used as the reference category analogous to the modelling with bribery as the item with the highest sensitivity assessment (see table 3). It can be seen that the assessment of nonresponse does not differ significantly between the bribery, infidelity and experience of violence items. This is particularly interesting as the sensitivity assessment between these three items differed significantly (see table 3). The corresponding analysis shows that the sensitivity of the infidelity item is rated the highest followed by the items on the experience of violence and bribery. The differences between the sensitivity assessments are significant, although the differences in the nonresponse assessments are not significant. This could be an indication that the hypothesized signalling effect of nonresponse is visible in the data. Nonresponse is not significantly more likely for items that produce a higher sensitivity rating, although this is initially expected from the significant coefficient on the sensitivity assessment in model 1.

Now that the sensitivity assessment and probability of nonresponse have been considered, the effect of various variables on intentional misreporting will be analysed as the third step. For this purpose, analyses are carried out with the central independent variable (sensitivity assessment) and other variables at the individual level. The multilevel regression models estimating the probability of misreporting are shown in table 5. A positive sign of a coefficient means that the independent variable increases the probability of intentional misreporting. The models in table 5 only include respondents who answered 'yes' to the previous question *'Would you answer this question?'*.³⁸

³⁸ In a separate model the analysis includes respondents who initially preferred to refuse to answer and these coefficients are used for comparison (see model 2 in table A5 in the appendix). These respondents were also asked to estimate how truthfully they would answer if they had to.

Table 5: Multilevel analysis of the probability of misreporting

	(1)	(2)	(3)	(4)
Sensitivity assessment	0.21***	0.21***	0.16***	0.16***
Violation of Anonymity (ref.: V. very unlikely)				
V. unlikely		0.24**	0.12	0.12
V. likely		0.36***	0.20*	0.19
V. very likely		0.36***	0.34**	0.32**
Violation of Anonymity # Sensitivity (ref.: V. very unlikely, no discomfort)				
V. unlikely # Sensitivity assessment			0.07**	0.07**
V. likely # Sensitivity assessment			0.09***	0.09***
V. very likely # Sensitivity assessment			0.03	0.02
Survey Experience (ref.: None)				
Rather rare				-0.03
Rather frequently				0.08
Frequently				-0.04
Education (ref: low)				
Medium				-0.41***
High				-0.36**
Gender (ref.: male)				
Female				0.03
Diverse				0.22
Age (in years)				0.01
IMC (ref.: not passed)				-0.18*
Constant	0.60***	0.38***	0.46***	0.60*
Sd (Cons)	1.62	1.61	1.60	1.59
Sd (Residual)	1.34	1.34	1.34	1.33
N (Level 2, Respondents)	1 593	1 593	1 593	1 593
N (Level 1, Vignettes)	5 909	5 909	5 909	5 909

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As in the previous analysis, higher values for the sensitivity parameter indicate a higher sensitivity assessment. The sensitivity assessment (H4) shows a significant influence on the assessment of intentional misreporting in all models. The positive sign indicates that respondents rate intentional misreporting as more likely with increasing sensitivity. In other words, the probability of intentional misreporting increases with the increase of assessed sensitivity even if the effect is controlled for individual characteristics of the respondents.³⁹ This statistic therefore confirms the validity of H4.

³⁹ The aforementioned holds true for the substitution of the sensitivity variable with the survey item variable. The additionally calculated models each show the same tendencies with regard to the influences of the other independent variables. The advantage of using the sensitivity estimate is a clearer interpretation of the interaction effect in model 3, which does not become significant when using the individual values of the survey items.

The second important variable for explaining the probability of misreporting is the anonymity assessment (H5) of the survey method used in the vignette. For a more straightforward interpretation of the interaction term (e.g., model 3) higher values of this variable represent a lower estimation of anonymity respectively a stronger violation of anonymity. The effect of violation of anonymity is assessed as an indicator since this variable has only four values. In model 2, the assessed violation of anonymity shows a significant and positive effect for all characteristics. As confidence in the anonymity of the responses decreases the probability of intentional misreporting increases. This is consistent with the postulated relationship in H5.

In model 3, the influences of sensitivity assessment and violation of anonymity are additionally added in the form of an interaction term (H6). This changes the amounts of the two main effects as well as their interpretation. The conditional main effect of the sensitivity assessment estimates the influence exclusively for respondents who consider a violation of anonymity to be very unlikely. Within this group of respondents, there is also a (to a certain extent isolated) reinforcing effect of the sensitivity assessment for intentional misreporting. The second conditional main effect for the assessment of the violation of anonymity is now also based on a different reference category. This coefficient applies to respondents who do not categorise the situation as sensitive. For these respondents, there is no significant difference between *"violation very unlikely"* and *"violation rather likely"*. The two higher values of the assessment of the violation of anonymity show significant differences with regard to the assessment of intentional misreporting. Both coefficients are significant and with a positive sign. This means that despite a very low sensitivity assessment by respondent's, greater concern about the anonymity of the responses leads to a higher assessment of the probability of intentional misreporting. Focusing on the interaction term in model 3, two out of three values are significant. Respondents who rate the violation of anonymity as either *"rather unlikely"* or *"rather likely"* show a higher probability of intentional misreporting with increasing sensitivity assessment. Interestingly, the third expression of the interaction term which stands for a *'very likely'* violation of anonymity and an increasing sensitivity assessment is not significant.

In order to further validate the calculated effects of the third model, control variables are added in model 4. Respondents' survey experience with the survey method described in the vignette is used as an indicator of whether respondents who participate in surveys more often

systematically misreport more or less often. This idea seems to be rejected based on model 4, as there are no significant coefficients and the signs do not show a consistent pattern.

The next control variable in model 4 is the level of education, which is divided into three levels. The two dummy variables for medium and high education show a significant difference to the reference category of low education. It is interesting to note that the coefficients show a stronger difference between low and medium education. Respondents with a medium level of education therefore rate their probability to intentional misreporting as the highest. Respondents with a high level of education also indicate higher degrees of intentional misreporting than respondents with a low level of education. A further model was used to test whether there is a significant difference between medium and high education. This distinction is not statistically significant (see model 1 in table A5 in the appendix).

The attention test in the form of an Instructional Manipulation Check (IMC) is significant in model 4 (see table 5). Respondents who do not successfully complete the IMC chose significantly lower values for their probability of intentional misreporting. Finally, considering the changes in the conditional main effects and the interaction effect between models 3 and 4, there is a change in the conditional main effect for the violation of anonymity. The coefficient for the answer option "*rather likely*" becomes insignificant in this model. The coefficient for a very likely violation of anonymity remains almost unchanged and is still clearly significant despite the additional variables. The situation is similar for the interaction effect, which does not differ significantly from model 3 in terms of coefficient size and significance.

⁴⁰ Overall, H6 is not rejected by these results.

Conclusion

This article aimed to explore factors driving the sensitivity assessment of respondents and the relationship between respondents' sensitivity rating and their response behaviour (potential nonresponse as well as intentional misreporting). The applied factorial survey design provided

⁴⁰ Finally, those respondents who initially preferred to refuse to answer are included in an additional model (see model 2 in table A5 in the appendix). Overall, the comparison with model 4 shows that most of the significant coefficients increase in magnitude with the inclusion of respondents who rated how they would respond if the preferred nonresponse is not possible. It should also be noted that the attention item (IMC) is insignificant in this model. This means that in this last model the distinction between respondents with regard to attentive reading of the instructions does not influence the probability of intentional misreporting.

insights into the relationship between sensitivity assessments and response behaviour without directly exposing respondents to a sensitive situation. Additionally, respondents assess sensitivity themselves without researchers determining which topics are sensitive. The results of this research are summarised in this chapter and critically evaluated in the final chapter. The conclusion is divided in two sections related to the dependent variables' sensitivity assessment and response behaviour.

Sensitivity assessment

Initially, the factors that could influence the respondents' sensitivity assessment were analysed (H1-H3). An ANOVA analysis and various multilevel regression analyses showed a significant influence of the survey item described in the vignette on the assessment of sensitivity. Therefore, H1 is not rejected.

Nevertheless, some surprising relationships emerged in the course of analysis that did not match previous expectations. The bribery item in the vignette was perceived by respondents as being less sensitive than expected based on the literature. The item asking about infidelity in the relationship was rated as the most sensitive. Furthermore, a separate calculation showed that the issue of sexual orientation is considered sensitive by members of the minority (non-heterosexual orientation) as well as members of the majority (heterosexual orientation). This does not correspond to the theoretical derivation that respondents who conform to a norm should evaluate less sensitivity. Summarising these three examples, it seems that a theoretical classification alone is not sufficient to assess the degree of sensitivity.

With regard to the mentioned presence of bystanders (H3) or interviewers (H2) there were initially no significant changes in the sensitivity assessments. Overall, H2 and H3 have to be rejected, because only very specific models showed some significant effects of these independent variables.

More detailed analysis reveals the relevance of these frameworks for specific topics. For instance, the mention of partners or family members and a question about infidelity led to a higher assessment of sensitivity. Other combinations of survey items and bystanders did not change the sensitivity assessment. This might be explained by referring to the study carried out by Aquilino, Wright and Supple (2000). In their study of the influence of bystanders on

responses to questions about alcohol and illicit drug use, the authors found that different people influenced response behaviour in different ways. If the bystander already knows the information this increases the probability of the respondent being honest. However, if the other person does not already know the relevant information the respondent not want the third party to find out (Aquilino et al., 2000, p. 864f). This study was related to the response behaviour of respondents. In the light of the results presented here, it can be assumed that the knowledge of bystanders might already has an effect on the sensitivity assessment. In the case of the question on infidelity in partnerships it can be assumed that partners or family members do not know the respondent's preliminary answer at least in the case of actual infidelity. At the same time the opposite effect (a lower sensitivity assessment due to the presence of partners or family members) was not found. In other words, there does not appear any advantage to the presence of bystanders in a survey but in one case there is a disadvantage (increased sensitivity assessment). It can therefore be confirmed that the presence of bystanders in interviews should be avoided at best. If this cannot be completely ruled out due to certain survey settings (e.g., online surveys) then at least a control question should be included in the questionnaire.

With regard to the presence of an interviewer there was a - rather surprising - deviation in the sensitivity assessment for the combination with the bribery item. When the vignette included both having an interviewer in a face-to-face interview and asking about active bribery respondents rated sensitivity lower. This was not expected on the basis of theoretical considerations. Two conclusions can be reached. On the one hand, the topic of bribery may be too abstract for the German population sample and therefore not lead to the expected processes (threat to disclose or social desirability bias). On the other hand, and related to this it could be an effect of the sample having less experience with bribery. A study on the specific topic of corruption might require a specific group of respondents who, for instance, tend to have more financial resources and opportunities to engage in acts of bribery. For instance, as Dickel and Graeff (2018) also did using a factorial survey, it would be conceivable to survey entrepreneurs.

Response behaviour

In the calculated models for the probability of nonresponse, the respondents' sensitivity assessment showed a clear relevance. Higher sensitivity assessments led to a higher probability of nonresponse. This initially emphasised the assumption that the nonresponse rate could be used as a measure of sensitivity. When the survey items of the vignettes were analysed separately, there were differences that were inconsistent with this linear relationship. The *sensitivity* of the questions on infidelity in partnerships, experience of violence and bribery was assessed differently. However, the assessment of *nonresponse* did not differ significantly between these items. This suggests that the nonresponse rate only partially reflects or indicates the sensitivity of survey data.

This impression was reinforced when the respondents' assessment of the anonymity was included in the estimate of nonresponse. It was seen that increasing concern about the anonymity of responses increased the probability of nonresponse. The nonresponse rate is therefore likely to reflect not only the sensitivity but also possible concerns of respondents regarding the anonymity of their responses. The nonresponse rate is therefore only suitable as an indicator for a sensitivity assessment if anonymity concerns can be addressed. In summary, it should be noted that the nonresponse rate should only be used as a clue of sensitivity and not as an operationalisation of it.

The third dependent variable of interest was the respondents' assessment of the probability of intentional misreporting. In these models, the sensitivity assessment of respondents (H4) showed a significant increase in the estimated probability of misreporting. This is in line with H4.

Furthermore, increasing concerns about the anonymity of responses (H5) increased the probability of intentional misreporting. If the anonymity was perceived to be at risk, both nonresponse and deliberate misreporting were evaluated as more likely. These results support H5.

In further analyses, the linear relationship between the assumption of a violation of anonymity and intentional misreporting turned out to be a weak interaction effect (H6). Assessing

anonymity combined with assessing sensitivity affects the estimated probability of intentional misreporting. H6 postulated this interaction. Worth mentioning was also the significant conditional main effect of the sensitivity assessment. Assuming that responses were very likely to remain anonymous there was still a significant relationship between increasing sensitivity assessment and a higher probability of intentional misreporting. These results suggest that even anonymity of respondents' answers is not sufficient to prevent deliberate misreporting regarding sensitive topics.

Discussion

A limitation of this study and of factorial surveys in general is the cognitive burden for respondents by the design compared to classical surveys with direct questions (Auspurg et al., 2014, p. 137).

There is also a focus on deliberate response bias. The literature suggests that unconscious and automated response behaviour can also lead to bias (Tourangeau et al., 2000, p. 14). However, this cannot be taken into account in an indirect survey design and would require respondents to have access to their potential unconscious sources of misreporting.

It was also not possible to assess the preliminary answer of respondents. Therefore, it is still unknown to which extent the sensitivity assessment is related to the preliminary answer.

Finally, the last limitation of this study is the use of direct questions (e.g., on demographic variables or on the respondents' subjective assessment of the anonymity). For these items, an unbiased response from the respondents had to be assumed in order to interpret the results.

The limitations of this study can also be fruitful for further research. A potential area for future research is the sensitivity assessment of further survey items. As mentioned above, the anonymity assessment or even demographic information could be influenced by sensitivity (bias). The results underline the relevance of specific examinations of sensitivity and its impact on respondents' response behaviour for even more subject areas. It would be particularly interesting to analyse survey items that have not yet been considered sensitive and are widely used. Social capital indicators could be one interesting example.

Finally, the investigation of additional factors that may moderate the association between sensitivity and intentional misreporting also deserve attention. Taking into account the response process described in the literature (Krosnick, 1991, p. 215) a more detailed consideration of respondents' motivation within the survey situation could be promising. Stronger motivation and increased respondent involvement may enhance the effects of sensitivity on response behaviour.

References

- Anduiza, E., & Galais, C. (2017). Answering Without Reading: IMCs and Strong Satisficing in Online Surveys. *International Journal of Public Opinion Research*, 29(3), 497-519. doi:10.1093/ijpor/edw007.
- Aquilino, W. S., Wright, D. L., & Supple, A. J. (2000). Response Effects Due to Bystander Presence in CASI and Paper-and-Pencil Surveys of Drug Use and Alcohol Use. *Substance Use & Misuse*, 35(6-8), 845-867. doi:10.3109/1082608000914842.
- Atzmüller, C., & Steiner, P. M. (2010). Experimental Vignette Studies in Survey Research. *Methodology*, 6(3), 128–138. doi:10.1027/1614-2241/a000014.
- Auspurg, K., Hinz, T., Liebig, S., & Sauer, C. (2014). The Factorial Survey as a Method for Measuring Sensitive Issues. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, & P. Sturgis, *Improving Survey Methods* (p. 137-149). New York: Routledge. doi:10.4324/9781315756288.
- Biemer, P., & Brown, G. (2005). Model-based Estimation of Drug Use Prevalence Using Item Count Data. *Journal of Official Statistics*, 21(2), 287–308.
- Blair, G., Coppock, A., & Moor, M. (2020). When to Worry about Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments. *American Political Science Review*, 114(4), 1297–1315. doi:10.1017/S0003055420000374.
- Blair, G., Imai, K., & Zhou, Y.-Y. (2015). Design and Analysis of the Randomized Response Technique. *Journal of the American Statistical Association*, 110(511), 1304-1319. doi:10.1080/01621459.2015.1050028.

- Böckenholt, U. (2014). Modeling motivated misreports to sensitive survey questions. *Psychometrika*, 79(3), 515-537. doi:10.1007/S11336-013-9390-9.
- Buchanan, T. (2000). Potential of the Internet for Personality Research. In M. H. Birnbaum, *Psychological experiments on the Internet* (p. 121-140). San Diego: Academic Press.
- Celhay, P., Meyer, B. D., & Mittag, N. (2022). *What Leads to Measurement Errors? Evidence from Reports of Program Participation in Three Surveys, IZA Discussion Papers*. Bonn: Institute of Labor Economics (IZA).
- Cobo, B., Casillo, E., López-Torrecillas, F., & Del Mar Rueda, M. (2021). Indirect questioning methods for sensitive survey questions: Modelling criminal behaviours among a prison population. *PLoS ONE*, 16(1), e0245550. doi:10.1371/journal.pone.0245550.
- Couper, M. P., Singer, E., Conrad, F. G., & Groves, R. M. (2008). Risk of Disclosure, Perceptions of Risk, and Concerns about Privacy and Confidentiality as Factors in Survey Participation. *Journal of Official Statistics*, 24(2), 255–275.
- Coutts, E., & Jann, B. (2011). Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT). *Sociological Methods & Research*, 40, 169-193.
- Czymara, C. S., & Schmidt-Catran, A. (2016). Wer ist in Deutschland willkommen? Eine Vignettenanalyse zur Akzeptanz von Einwanderern. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 68, 193-227. doi:10.1007/s11577-016-0361-x.
- Delios, A., Malesky, E. J., Yu, S., & Riddler, G. (2024). Methodological errors in corruption research. Recommendations for future research. *Journal of International Business Studies*, 55, 235-251. doi:10.1057/s41267-023-00637-8.
- DeMaio, T. J. (1984). Social Desirability and Survey Measurement. A Review. In C. F. Turner, & E. Martin, *Surveying subjective Phenomena* (p. 257-282). New York: Russell Sage Foundation.
- Dickel, P., & Graeff, P. (2018). Entrepreneurs' propensity for corruption: A vignette-based factorial survey. *Journal of Business Research*, 89, 77-86. doi:10.3109/10826080009148424.

- Diop, A., Le, K. T., & Traugott, M. (2015). Third-Party Presence Effect with Propensity Score Matching. *Journal of Survey Statistics and Methodology*, 3, 193-215. doi:10.1093/jssam/smv005.
- Droitcour, J. A., Caspar, R., Hubbard, M., Parsely, T., Visscher, W., & Ezzati, T. (1991). The item count technique as a method of indirect questioning: a review of its development and a case study application. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, & S. Sudman, *Measurement Errors in Surveys* (p. 185-210). New York: Wiley.
- Druckman, J. N., Gilli, M., Klar, S., & Robison, J. (2015). Measuring Drug and Alcohol Use Among College Student-Athletes. *Social Science Quarterly*, 96, 369-380.
- Dülmer, H. (2016). The Factorial Survey: Design Selection and its Impact on Reliability and Internal Validity. *Sociological Methods & Research*, 1-44. doi: 10.1177/0049124115582269.
- Europass by European Union. (no date). *Compare national qualifications frameworks across Europe. Information about national qualifications frameworks and qualification levels.* [https://europa.eu/europass/en/compare-qualifications?field_eqf_selection_target_id=undefined&field_location_selection_target_id\[6068\]=6068&field_location_selection_target_id\[0\]=0](https://europa.eu/europass/en/compare-qualifications?field_eqf_selection_target_id=undefined&field_location_selection_target_id[6068]=6068&field_location_selection_target_id[0]=0)
- European Union Agency for Fundamental Rights. (2020). *A long way to go for LGBTI equality.* Von Publications Office of the European Union: <https://data.europa.eu/doi/10.2811/7746>.
- Glynn, A. N. (2013). What can we learn with statistical truth serum? Design and Analysis of the List Experiment. *The Public Opinion Quarterly*, 77, 159-172.
- Gnambs, T., & Kaspar, K. (2015). Disclosure of sensitive behaviors across self-administered survey modes: a meta-analysis. *Behavior Research Methods*, 47, 1237-1259. doi:10.3758/s13428-014-0533-4.
- Haber, N., Harling, G., Cohen, J., Mutevedzi, T., Tanser, F., Gareta, D., . . . Fink, G. (2018). List randomization for eliciting HIV status and sexual behaviors in rural KwaZulu-Natal, South Africa: a randomized experiment using known true values for validation. *BMC Medical Research Methodology*, 18(46), 1-12. doi:10.1186/s12874-018-0507-9.

- Hinsley, A., Keane, A., St. John, F. A., Ibbett, H., & Nuno, A. (2019). Asking sensitive questions using the unmatched count technique: Applications and guidelines for conservation. *Methods in Ecology and Evolution*, 10, 308-319. doi:10.1111/2041-210X.13137.
- Holbrook, A., & Krosnick, J. A. (2010). Social Desirability Bias in Voter Turnout Reports. Tests using the Item Count Technique. *Public Opinion Quarterly*, 74(1), 37-67.
- Holtgraves, T. (2004). Social Desirability and Self-Reports: Testing Models of Socially Desirable Responding. *Personality and Social Psychology Bulletin*, 30(2), 161-172. doi:10.1177/0146167203259930
- Hox, J. J. (2010). *Multilevel Analysis. Techniques and Applications* (2nd edition). New York: Routledge.
- Huber-Krum, S., Karadon, D., Kurutas, S., Rohr, J., Baykal, S. S., Okcuoglu, B. A., . . . Shah, I. (2020). Estimating abortion prevalence and understanding perspectives of community leaders and providers: Results from a mixed-method study in Istanbul, Turkey. *Women's Health*, 16, 1-13. doi:10.1177/1745506520953353.
- Ibbett, H., Jones, J. P., & St John, F. A. (2021). Asking sensitive questions in conservation using Randomised Response Technique. *Biological Conservation*, 260, 109191.
- Jasso, G. (2006). Factorial Survey Methods for Studying Beliefs and Judgments. *Sociological Methods & Research*, 34(3), 334-423. doi:10.1177/004912410528312.
- Kantar, Public Division. (2019). *(Strukturelle) Diskriminierung. Eine Studie von Kantar, Public Division im Auftrag des Presse- und Informationsamtes der Bundesregierung (BPA)*.
- Kissling, A., & Jackson, H. M. (2022). Estimating Prevalence of Abortion Using List Experiments: Findings from a Survey of Women in Delaware and Maryland. *Women's Health Issues*, 32(1), 33-40. doi:10.1016/j.whi.2021.08.003.
- Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social Desirability Bias in CATI, IVR, and Web Surveys. The effects of mode and question sensitivity. *Public Opinion Quarterly*, 72(5), 847-864. doi:10.1093/poq/nfn063.
- Krosnick, J. A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, 5, 213-236.

- Krosnick, J. A. (1999). Survey Research. *Annual Review of Psychology*, 50, 537-567.
- Krosnick, J. A. (2018). Confidentiality, Privacy, and Anonymity. In D. L. Vannette, & J. A. Krosnick, *The Palgrave Handbook of Survey Research* (p. 347-350). palgrave macmillan. doi:10.1007/978-3-319-54395-6_58.
- Krumpal, I. (2012). Estimating the prevalence of xenophobia and anti-Semitism in Germany: A comparison of randomized response and direct questioning. *Social Science Research*, 41, 1387–1403.
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Qual Quant*, 47, 2025–2047. doi:10.1007/s11135-011-9640-9.
- Krumpal, I., & Voss, T. (2020). Sensitive Questions and Trust: Explaining Respondents' Behavior in Randomized Response Surveys. *SAGE Open*, 10(3). doi:10.1177/2158244020936223.
- Kuhn, P. M., & Vivyan, N. (2021). The Misreporting Trade-Off Between List Experiments and Direct Questions in Practice: Partition Validation Evidence from Two Countries. *Political Analysis*, 30(3), 381- 402. doi:10.1017/pan.2021.10.
- Kuklinski, J. H., Cobb, M. D., & Gilens, M. (1997). Racial Attitudes and the 'New South'. *Journal of Politics*, 59(2), 323-49.
- Lee, R. M., & Renzetti, C. M. (1990). The Problems of Researching Sensitive Topics: An Overview and Introduction. *American Behavioral Scientist*, 33(5), 510-528. doi:10.1177/0002764290033005002.
- Lelkes, Y., Krosnick, J. A., Marx, D. M., Judd, C. M., & Park, B. (2012). Complete anonymity compromises the accuracy of self-reports. *Journal of Experimental Social Psychology*, 48, 1291-1299. doi:10.1016/j.jesp.2012.07.002.
- Lensvelt-Mulders, G. J., Hox, J. J., van der Heijden, P. G., & Maas, C. J. (2005). Meta-Analysis of Randomized Response Research. Thirty-Five Years of Validation. *Sociological Methods & Research*, 33, 319-348.
- Lépine, A., Treibich, C., Ndour, C. T., Gueye, K., & Vickerman, P. (2020). HIV infection risk and condom use among sex workers in Senegal: evidence from the list experiment method. *Health Policy and Planning*, 35, 408-415. doi:10.1093/heapol/czz155.

- Miller, J. D. (1984). *A new survey technique for studying deviant behavior*. The George Washington University.: PhD diss.
- Mneimneh, Z. M., Tourangeau, R., Pennell, B.-E., Heeringa, S. G., & Elliott, M. R. (2015). Cultural Variations in the Effect of Interview Privacy and the Need for Social Conformity on Reporting Sensitive Information. *Journal of Official Statistics*, 31(4), 673-697. doi:10.1515/JOS-2015-0040.
- Moseson, H., Gerdtz, C., Dehlendorf, C., Hiatt, R. A., & Vittinghoff, E. (2017). Multivariable regression analysis of list experiment data on abortion: results from a large, randomlyselected population based study in Liberia. *Population Health Metrics*, 15(40), 1-8. doi:10.1186/s12963-017-0157-x.
- Mutz, D. C. (2011). *Population-Based Survey Experiments*. Princeton: Princeton University Press. doi:10.1515/9781400840489.
- Ong, A. D., & Weiss, D. J. (2000). The Impact of Anonymity on Responses to Sensitive Questions. *Journal of Applied Social Psychology*, 30(8), 1691-1708.
- Preisendörfer, P., & Wolter, F. (2014). Who Is Telling the Truth? A Validation Study on Determinants of Response Behavior in Surveys. *Public Opinion Quarterly*, 78(1), 126146. doi:10.1093/poq/nft079.
- Rasinski, K. A., Willis, G. B., Baldwin, A. K., Yeh, W., & Lee, L. (1999). Methods of Data Collection, Perceptions of Risks and Losses, and Motivation to Give Truthful Answers to Sensitive Survey Questions. *Applied cognitive Psychology*, 13, 465-484.
- Richards, L. (2017). Using Survey Methods to Research Corruption. In A. K. Schwickerath, A. Varraich, & L.-L. Smith, *How to research corruption? Conference Proceedings: Interdisciplinary Corruption Research Forum, June 2016* (p. 4-12).
- Roe-Sepowitz, D., Bontrager, S., Pickett, J. T., & Kosloski, A. E. (2019). Estimating the sex buying behavior of adult males in the United States: List experiment and direct question estimates. *Journal of Criminal Justice*, 63, 41-48. doi:10.1016/j.jcrimjus.2019.04.005.
- Rosenfeld, B., Imai, K., & Shapiro, J. N. (2016). An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions. *American Journal of Political Science*, 60(3), 783-802.

- Schlüter, E., & Schmidt, P. (2010). Special issue: Survey experiments. *Methodology European Journal of Research Methods for the Behavioral and Social Sciences*, 6, 93-54.
- Simpser, A. (2017). When Do Sensitive Survey Questions Elicit Truthful Answers? Theory and Evidence with Application to the RRT and the List Experiment. doi:10.2139/ssr.
- Singer, E., & Ye, C. (2013). The Use and Effects of Incentives on in Surveys. *The ANNALS of the American Academy of Political and Social Science*, 645, 112-141.
- Singer, E., Mathiowetz, N. A., & Couper, M. P. (1993). The impact of privacy and confidentiality concerns on survey participation: The case of the 1990 U.S. census. *Public Opinion Quarterly*, 57, 465-482. doi:10.1086/269391.
- Smith, T. W. (2014). Collecting Survey Data on Sensitive Topics: Sexual Behavior. In T. P. Johnson, *Handbook of Health Survey Methods* (p. 1139-1171). New Jersey: John Wiley & Sons. doi:10.1002/9781118594629.
- Tourangeau, R. (2018). Confidentiality, Privacy, and Anonymity. In D. L. Vannette, & J. A. Krosnick, *The Palgrave Handbook of Survey Research* (p. 501-507). palgrave macmillan. doi:10.1007/978-3-319-54395-6_58.
- Tourangeau, R., & Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin*, 133(5), 859–883. doi:10.1037/0033-2909.133.5.859.
- Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Trappmann, M., Krumpal, I., Kirchner, A., & Jann, B. (2014). Item Sum: A new Technique for Asking quantitative sensitive Questions. *Journal of Survey Statistics and Methodology*, 2, 58-77. doi:10.1093/jssam/smt019.
- Treischl, E., & Wolbring, T. (2022). The Past, Present and Future of Factorial Survey Experiments: A Review for the Social Sciences. *methods, data, analyses*, 16(2), 141170. doi:10.12758/mda.2021.07.
- Vannette, D. L. (2018). Best Practices for Survey Research. In D. L. Vannette, & J. A. Krosnick, *The Palgrave Handbook of Survey Research* (p. 331-343). palgrave macmillan. doi:10.1007/978-3-319-54395-6_58.

- Vannette, D. L., & Krosnick, J. A. (2014). Answering Questions: A Comparison of Survey Satisficing and Mindlessness. In A. Ie, C. T. Ngnoumen, & E. J. Langer, *The Wiley Blackwell Handbook of Mindfulness* (p. 312-327). John Wiley & Sons. doi:10.1002/9781118294895.ch17.
- Vriesema, C. C., & Gehlbach, H. (2021). Assessing Survey Satisficing: The Impact of Unmotivated Questionnaire Responding on Data Quality. *Educational Researcher*, 50(9), 618-627. doi:10.3102/0013189X211040054.
- Wallander, L. (2009). 25 years of factorial surveys in sociology: A review. *Social Science Research* (38), 505-520. doi:10.1016/j.ssresearch.2009.03.004.
- Warner, S. L. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, 60(309), 63-69.
- Wolter, F. (2019). A New Version of the Item Count Technique for Asking Sensitive Questions: Testing the Performance of the Person Count Technique. *methods, data, analyses*, 13(1), 169-192. doi:10.12758/mda.2018.04.
- Wolter, F., & Preisendörfer, P. (2013). Asking Sensitive Questions: An Evaluation of the Randomized Response Technique Versus Direct Questioning Using Individual Validation Data. *Sociological Methods & Research*, 42(3), 321-353. doi:10.1177/0049124113500474.
- Yan, T. (2021). Consequences of Asking Sensitive Questions in Surveys. *Annual Review of Statistics and Its Application*, 8, 109–27. doi:10.1146/annurev-statistics-040720033353

Appendix

Table A1: Overview of vignette variables

Dimensions describing the scenario:	Levels (values) for the dimensions (situational characteristics)	Wording in the vignette scenario
Survey item	Age (reference)	How old are you? (in years)
	Religion	What religion do you feel you belong to? (Christianity / Islam / Judaism / Hinduism / Buddhism / Other / None / Prefer not to say)
	Sexual Orientation	Which of the following describes your sexual orientation (most accurately)? (Heterosexual / Gay / Lesbian / Bisexual / Queer / Asexual / Pansexual / Ambiguous or insecure / Other Orientation / Prefer not to say)
	Experience of violence	Have you been the victim of physical violence by anyone in the past 12 months? (Yes / No / Prefer not to say.)
	Adultery	Have you ever been unfaithful to someone in a relationship? (Yes, once. / Yes, multiple times with the same person. / Yes, multiple times with different people. / No, but I've been tempted. / No, never.)
	Bribery	In the last 5 years, have you offered someone money to do something for you that no one else should know about? (Yes / No / Prefer not to say)
Survey mode	Web survey (reference)	You are taking part in an online survey.
	Telephone survey	You are taking part in a telephone survey.
	Face-to-Face survey	You are taking part in a Face-to-Face survey
Presence of bystanders	None (reference)	No other people are present at the time.
	Partner	Meanwhile, your partner is in the same room as you.
	Family member	Meanwhile, a family member (not your partner) is in the same room as you.

Table A2: Formulation of the items for anonymity assessment and the instrumental manipulation check

How likely do you think it is that your answers will remain anonymous (meaning only known to researchers)?			
... in online surveys?			
☐ Very likely	☐ Rather likely	☐ Rather unlikely	☐ Very unlikely
... in telephone surveys?			
☐ Very likely	☐ Rather likely	☐ Rather unlikely	☐ Very unlikely
... in personal interviews?			
☐ Very likely	☐ Rather likely	☐ Rather unlikely	☐ Very unlikely

How many telephones do you have in your household? (IMC)	
Select 0 to show that you are attentive during the survey	
☐	0
☐	1
☐	2
☐	3 or more

How often in your life have you...			
... participated in online surveys?			
☐ Never	☐ Rather seldom	☐ Rather frequently	☐ Frequently
... participated in telephone surveys?			
☐ Never	☐ Rather seldom	☐ Rather frequently	☐ Frequently
... participated in personal interviews?			
☐ Never	☐ Rather seldom	☐ Rather frequently	☐ Frequently

Note: This English version is translated from the original German questionnaire. These items were presented on separate pages in the original questionnaire.

Table A3: Pairwise correlations between vignette dimensions and Chi²
Tests between vignette dimensions

Pearson's r	Survey mode	Presence of bystanders
Presence of bystanders	0.0057 (0.6436)	-
Survey item	-0.0022 (0.8598)	0.0013 (0.9156)

N = 6 568, p-values in parentheses

Levels of 'Survey mode'	Levels of 'Presence of bystanders'		
	None	Partner	Family member
Web survey	735	725	723
Telephone survey	731	726	722
Personal interview	727	732	739
Pearson's chi ² = 0.2408 (p = 0.993)			

Levels of 'Survey mode'	Levels of 'Survey item'					
	Age	Religion	Sexual orientation	Infidelity	Experience of violence	Bribery
Web survey	360	368	368	358	366	363
Telephone survey	363	358	368	367	359	364
Personal interview	371	369	365	363	365	364
Pearson's chi ² = 0.5016 (p = 1.000)						

Levels of 'Survey mode'	Levels of 'Survey item'					
	Age	Religion	Sexual orientation	Infidelity	Experience of violence	Bribery
Web survey	365	368	369	365	359	367
Telephone survey	365	365	367	358	370	358
Personal interview	364	362	365	365	361	367
Pearson's $\chi^2 = 0.4729$ ($p = 1.000$)						

Table A4: Additional multilevel models for sensitivity assessment

	(1)	(2)	(3)	(4)	(5)	(6)
Survey Item (ref.: Age)						-0.95***
Religion	0.10	0.03	0.18	0.18	-0.09	-1.04***
Sexual orientation	0.62***	0.77***	0.76***	0.76***	0.63***	-0.32*
Infidelity	1.90***	1.82***	1.67***	1.67***	1.45***	0.51**
Experience of violence	1.60***	1.55***	1.64***	1.65***	1.54***	0.59***
Bribery	1.02***	1.02***	1.31***	1.31***	0.95***	
Sexual Orientation (ref.: heterosexual)						
Non- heterosexual	0.05					
Survey Item # Sexual Orientation (ref.: age, heterosexual)						
Religion # non-heterosexual	-0.31					
Sexual orientation # non-heterosexual	0.60**					
Infidelity # non-heterosexual	-0.33					
Experience of violence # non-heterosexual	-0.24					
Bribery # non-heterosexual	0.00					
Survey Mode (ref.: Web Survey)						
Telephone survey	0.11			0.16	0.12	0.12
Personal interview	0.10			0.24	0.10	0.10
Presence of bystanders (ref.: None)						
Partner	0.02	0.01	0.01	0.01	-0.10	-0.02
Family member	0.09	0.08	0.08	0.07	-0.18	-0.03
Presence of interviewer (ref.: None)						
Interviewer (in personal or one the phone)		0.11	0.20			
Survey Item # Presence of bystanders (ref.: age or bribery, no bystanders)						
Age # Partner						-0.08
Age # Family member						-0.15
Religion # Partner					0.18	0.09
Religion # Family member					0.19	0.04
Sexual orientation # Partner					0.05	-0.03
Sexual orientation # Family member					0.38	0.23
Infidelity # Partner					0.56*	0.48*
Infidelity # Family member					0.54*	0.39
Experience of violence # Partner					-0.21	-0.29
Experience of violence # Family member					0.25	0.10
Bribery # Partner					0.08	
Bribery # Family member					0.15	
Survey Item # Interviewer (ref.: age, no interviewer)						
Religion # Interviewer			-0.22			
Sexual orientation # Interviewer			0.03			
Infidelity # Interviewer			0.23			
Experience of violence # Interviewer			-0.15			

Bribery # Interviewer	-0.43*					
Survey Item # Survey mode (ref.: age, web survey)						
Religion # Telephone survey	-0.21					
Religion # Personal interview	-0.22					
Sexual orientation # Telephone survey	0.20					
Sexual orientation # Personal interview	-0.15					
Infidelity # Telephone survey	0.19					
Infidelity # Personal interview	0.26					
Experience of violence # Telephone survey	-0.04					
Experience of violence # Personal interview	-0.26					
Bribery # Telephone survey	-0.40					
Bribery # Personal interview	-0.47*					
Constant	1.12***	1.13***	1.07***	1.07***	1.25***	2.20***
Sd (Cons)	2.00	2.00	2.00	2.00	2.00	2.00
Sd (Residual)	1.96	1.97	1.96	1.96	1.96	1.96
N (Level 2, Respondents)	1 640	1 640	1 640	1 640	1 640	1 640
N (Level 1, Vignettes)	6 560	6 560	6 560	6 560	6 560	6 560

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A5: Additional multilevel models for the probability of misreporting

	(1)	(2)
Sensitivity assessment	0.16***	0.18***
Violation of Anonymity (ref.: V. very unlikely)		
V. unlikely	0.12	0.06
V. likely	0.19	0.11
V. very likely	0.32**	0.29*
Violation of Anonymity # Sensitivity (ref.: V. very unlikely, no discomfort)		
V. unlikely # Sensitivity assessment	0.07**	0.13***
V. likely # Sensitivity assessment	0.09***	0.17***
V. very likely # Sensitivity assessment	0.02	0.05
Survey Experience (ref.: None)		
Rather rare	-0.03	-0.01
Rather frequently	0.09	0.01
Frequently	-0.04	-0.09
Education		
Low	0.36**	
Medium	-0.05	-0.36**
High		-0.37**
Gender (ref.: male)		
Female	0.00	0.07
Diverse	0.22	0.01
Age (in years)	0.01	0.01
IMC (ref.: not passed)	-0.18*	-0.18
Constant	0.24	0.59*
Sd (Cons)	1.59	1.62
Sd (Residual)	1.34	1.55
N (Level 2, Respondents)	1 593	1 615
N (Level 1, Vignettes)	5 909	6 456

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5 DECLARATION ON INDEPENDENT WORK

Ich erkläre hiermit, dass ich meine Doktorarbeit „The Measurement of Sensitivity“ selbstständig und ohne fremde Hilfe angefertigt habe und dass ich als Koautorin maßgeblich zu den weiteren Fachartikeln beigetragen habe.

Alle von anderen Autoren wörtlich übernommenen Stellen, wie auch die sich an die Gedanken anderer Autoren eng anlehnenden Ausführungen der aufgeführten Beiträge wurden besonders gekennzeichnet und die Quellen nach den mir angegebenen Richtlinien zitiert.

Kiel, den 10.07.2024

Ort, Datum

F. Dunkelmann

Unterschrift F. Dunkelmann

6 DECLARATION ABOUT THE CO-AUTHORSHIP AND THE DFG-GUIDELINES

The book chapter titled: "Vignette studies" was written in co-authorship between Prof. Dr. Peter Graeff and Franziska Dunkelmann.

Hereby, we declare that both authors have participated in composing the article.

Furthermore, we confirm that the DFG *Guidelines for Safeguarding Good Research Practice* were followed.

Kiel, den 19.06.24

Ort, Datum



Unterschrift P. Graeff (Ko-Autor)

Kiel, den 10.07.2024

Ort, Datum



Unterschrift F. Dunkelmann (Doktorandin)