

SDM Survey Guidelines

Frageformulierung

Timo Lenzner & Natalja Menold

Zusammenfassung

Dieser Beitrag gibt einen Überblick über grundsätzliche Regeln, die bei der Formulierung von Survey-Fragen berücksichtigt werden sollten. Die Gliederung des Kapitels orientiert sich dabei am kognitiven Antwortprozess, den Befragte bei der Beantwortung von Survey-Fragen durchlaufen und der insgesamt vier Aufgaben umfasst: Verstehen des Fragetextes, Informationsgewinnung, Urteilsbildung und Formatierung der Antwort. Anhand von praktischen Beispielen wird illustriert, wie man mit Hilfe der Frageformulierung Befragten die Bewältigung dieser vier Aufgaben erleichtern kann.

Zitierung

Lenzner, Timo und Menold, Natalja (2015). Frageformulierung. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (SDM Survey Guidelines). DOI: 10.15465/sdm-sg_017

1. Welche Anforderungen müssen die Fragenformulierungen in Surveys erfüllen?

Eine Survey-Frage sollte von allen Befragten in der vom Fragebogenentwickler intendierten Weise interpretiert werden und möglichst leicht zu verstehen und zu beantworten sein. Dies ist eine Voraussetzung für eine fehlerfreie Datenerhebung. Der Frageformulierung kommt hierbei eine entscheidende Rolle zu, da sie einen direkten Einfluss darauf hat, wie leicht oder schwer es Befragten fällt, eine Frage in der von dem Forscher intendierten Weise zu verstehen und zu beantworten.

Beim Beantworten einer Survey-Frage müssen Befragte insgesamt vier kognitive Aufgaben bewältigen (Tourangeau, 1984; Tourangeau, Rips, & Rasinski, 2000):

- (1) die Frage verstehen und ihre Intention ableiten (*comprehension*),
- (2) relevante Informationen zum Beantworten der Frage aus dem Gedächtnis abrufen (*retrieval*),
- (3) auf Basis dieser Informationen ein Urteil bilden (*judgment*) und
- (4) dieses Urteil mit Hilfe eines Antwortformats zum Ausdruck bringen (*response*).

Je nachdem wie eine Survey-Frage formuliert ist, kann jede dieser Aufgaben mehr oder weniger schwierig zu bewältigen sein. Unpräzise Begriffe oder eine komplexe Satzstruktur erschweren beispielsweise das Verstehen einer Frage (Tourangeau et al., 2000). Fragt man Befragte nach Ereignissen, die schon lange Zeit zurückliegen, fällt es Ihnen gewöhnlich schwer, die relevanten Informationen zur Beantwortung der Frage aus dem Gedächtnis abzurufen und/oder auf Basis dieser Informationen ein Urteil zu bilden (Cannell, Miller, & Oksenberg, 1981). Antwortkategorien, die nicht eindeutig und überschneidungsfrei sind, erschweren es den Befragten, ihre Antwort mit Hilfe des vorgegebenen Antwortformats auszudrücken. Eine suboptimale Formulierung von Survey-Fragen kann daher die Fragenbeantwortung erschweren und die Genauigkeit der Antworten beeinträchtigen, z. B. können die Antworten aufgrund der Verständnisprobleme derart verzerrt werden, dass die intendierten Inhalte damit nicht abgebildet werden können.

Dieser Beitrag gibt Hinweise darauf, wie Survey-Fragen formuliert sein sollten, damit sie den kognitiven Aufwand bei der Beantwortung möglichst gering halten und eine möglichst fehlerfreie Datenerhebung ermöglichen. Die Gliederung des Kapitels orientiert sich dabei an den vier Stufen des kognitiven Antwortprozesses.

2. Wie kann man die Verständlichkeit von Survey-Fragen verbessern?

In der Umfrageforschung gilt es als allgemein anerkannt, dass Fragen so formuliert sein sollten, dass sie für die Befragten leicht zu verstehen sind und dass alle Befragten die Fragen auch so verstehen, wie sie vom Forscher intendiert sind (Fowler, 1992). Um diese beiden Ziele zu erreichen, sollte man unterschiedliche Aspekte der Frageformulierung beachten.

2.1 Vermeiden von unbekanntem und wenig geläufigen Begriffen

Unbekannte und wenig geläufige Begriffe bergen die Gefahr, dass sie von den Befragten überhaupt nicht oder unterschiedlich verstanden werden. Wenn Befragte einen Begriff nicht verstehen, so können sie die Frage kaum sinnvoll beantworten. Wenn zwei oder mehr Befragte eine Frage unterschiedlich

interpretieren, so unterscheiden sich ihre Antworten ggf. nicht aufgrund von verschiedenen Meinungen, sondern aufgrund des unterschiedlichen Fragenverständnisses.

Daher sollten unbekannte und wenig geläufige Begriffe wie Fremdwörter, Fachbegriffe, Abkürzungen, Akronyme und seltene Begriffe in Survey-Fragen vermieden werden (Graesser et al., 2006; Lenzner, 2011). Empfehlenswert ist es daher zum einen, generell alle Wörter in Survey-Fragen auszuschreiben und Akronyme zu vermeiden (z.B. „Atomkraftwerk“ statt „AKW“ oder „Erneuerbare-Energien-Gesetz“ statt „EEG“). Zum anderen sollte anhand eines (Online-)Wörterbuchs überprüft werden, ob es für die zentralen Begriffe in Survey-Fragen Synonyme gibt, die geläufiger und damit einfacher zu verstehen sind. Ist dies der Fall, so sollte man die weniger geläufigen Begriffe durch diese Synonyme ersetzen. Hier zwei Beispiele:

- (1) „Inwieweit stimmen Sie der folgenden Aussage zu?
Die sozialen *Diskrepanzen* (besser: *Unterschiede*) in Deutschland bleiben in Zukunft sicherlich bestehen.“
- (2) „Wie häufig kam es in den letzten vier Wochen vor, dass Sie *somatische* (besser: *körperliche*) Beschwerden hatten?“

Gibt es für wenig geläufige Begriffe keine geläufigeren Synonyme (3a), so sollte man die wenig geläufigen Begriffe definieren (3b). Hier ein Beispiel aus Porst (2008):

- (3a) „Was glauben Sie: In welchem Alter beginnt bei Männern normalerweise die *Andropause*?“
- (3b) „Mit dem Begriff *Andropause* umschreibt man mit dem Altern verbundene hormonelle Veränderungen beim Mann, die sich auf das Gefühls- und Sexualleben auswirken können. Vergleichbar ist dieser Prozess der Menopause bei Frauen, also den so genannten Wechseljahren. Was glauben Sie: In welchem Alter beginnt bei Männern normalerweise die *Andropause*?“

Welche Begriffe und Formulierungen leicht und eindeutig zu verstehen sind, hängt selbstverständlich von der Zielgruppe ab. Würde man beispielsweise Ärzte zum Umgang mit Patienten mit einer bestimmten Erkrankung befragen, so wäre es umständlich und im Sinne der Verständigung unökonomisch, Begriffe zu vermeiden (Dillman, Smyth, & Christian, 2009). In einem solchen Fall können Fachbegriffe genutzt werden bzw. wäre zu prüfen, inwieweit ein Begriff unter Fachexperten geläufig ist.

2.2 Vermeiden von unpräzisen relativen Bezeichnungen

Unpräzise relative Bezeichnungen, wie z.B. „oft“, „kürzlich“ oder „erheblich“ sind insofern problematisch, als dass sie sich implizit auf ein zugrundeliegendes Kontinuum beziehen, ihre genaue Lage auf diesem Kontinuum jedoch nicht eindeutig ist (Graesser et al., 2006; Lenzner, 2011). Daher können sie auf verschiedene Arten interpretiert werden, was die Gefahr erhöht, dass Befragte die Begriffe nicht in der Weise verstehen, wie sie vom Fragebogenentwickler beabsichtigt sind. Hier ein Beispiel:

- (4a) „Haben Sie *kürzlich* einen oder mehrere Ärzte aufgesucht?“
Wenn Befragte mit „ja“ antworten, so wären sie in einer weiteren Frage nach der Anzahl der Arztbesuche zu fragen.

Bei dieser Frage ist davon auszugehen, dass sich die Befragten in ihrer Interpretation des Begriffs „kürzlich“ sehr deutlich unterscheiden. Die einen werden an die letzten Tage, die anderen an die letzten Monate denken. Neben dem Problem, dass unterschiedliche Befragte den Begriff auf unterschiedliche Weise interpretieren, besteht hier außerdem die Gefahr, dass Befragten die Mehrdeutigkeit des Begriffs auffällt und ihnen die Beantwortung erschwert (Lenzner, 2012). So können sich Befragte z.B. wundern,

welcher Zeitraum mit „kürzlich“ gemeint ist und wenn sie sich nicht für einen bestimmten Zeitraum entscheiden können, die Frage unbeantwortet lassen. Daher sollten solche unpräzisen relativen Begriffe in Survey-Fragen (vor allem im Fragetext) vermieden und durch konkrete Begriffe ersetzt werden. Im oben aufgeführten Beispiel könnte eine Revision der Frage wie folgt aussehen:

- (4b) „Haben Sie in den *letzten vier Wochen* einen oder mehrere Ärzte aufgesucht?“
Wenn Befragte mit „ja“ antworten, so wären sie in einer weiteren Frage nach der Anzahl der Arztbesuche zu fragen.

Mehr zu solchen Fragen, von deren Beantwortung abhängig ist, ob weitere Fragen gestellt werden (sog. Filterfragen) siehe im SDM Survey Guidelines Artikel „Schriftlich-postalische Befragung“ Menold (2015).

2.3 Vermeiden von abstrakten und mehrdeutigen Begriffen

Abstrakte Begriffe kann man sehr leicht daran erkennen, dass sie mehrere spezifischere und konkretere Begriffe miteinschließen. Der Begriff „kulturelle Veranstaltungen“ in Frage (5a) schließt beispielsweise die konkreteren Begriffe „Theateraufführungen“, „Opern“, aber auch „Pop-Konzerte“ und „Rock Festivals“ mit ein.

- (5a) „Wie häufig besuchen Sie in Ihrer Freizeit *kulturelle Veranstaltungen*?“

Hier können manche Befragte nur an Ereignisse wie Theateraufführungen und klassische Konzerte denken, andere Befragte jedoch nur (oder auch zusätzlich) an Pop-Konzerte und Poetry Slams. Zum anderen kann es den Befragten aufgrund der Mehrdeutigkeit des Begriffs „kulturelle Veranstaltungen“ schwer fallen, sich für eine Interpretation des Begriffs zu entscheiden, so dass sie die Frage dann unbeantwortet lassen können (Lenzner, Kaczmirek, & Galesic, 2011; Lenzner, 2012). Um die Verständlichkeit von Fragen zu verbessern, sollten daher immer möglichst konkrete und spezifische Begriffe an Stelle abstrakter Begriffe verwendet werden. Alternativ könnten „kulturelle Veranstaltungen“ als Sammelbegriff mit drei Beispielen verdeutlicht werden. Die Frage oben könnte beispielsweise wie folgt überarbeitet werden:

- (5b) „Wie häufig besuchen Sie in Ihrer Freizeit *kulturelle Veranstaltungen, wie z.B. Konzerte, Lesungen oder Theateraufführungen*?“

Eine weitere Vereinfachung der Beantwortung kann darin bestehen, unterschiedliche Veranstaltungen aufzulisten und diese Liste zur Bewertung vorzulegen.

2.4 Vermeiden von komplexen Satzstrukturen

Die syntaktische Struktur von Survey-Fragen sollte möglichst geringe Komplexität aufweisen, damit sie die kognitive Kapazität der Befragten nicht überlastet. Beispielsweise kann man die grammatikalische Komplexität von Fragen verringern, indem man Passivsätze (6a) in Aktivsätze (6b) überführt (Lenzner, Kaczmirek, & Lenzner, 2010):

- (6a) „Inwieweit stimmen Sie der folgenden Aussage zu?
Zu viel Geld *wird* vom Staat *ausgegeben*, um Zuwanderer zu unterstützen.“
- (6b) Der Staat *gibt* zu viel Geld *aus*, um Zuwanderer zu unterstützen.“

Vermeiden sollte man darüber hinaus auch die Verwendung von Nominalisierungen, d.h. von Verben, die in Substantive transformiert wurden („erwerben“ → „Erwerb“):

- (7a) „Inwieweit stimmen Sie der folgenden Aussage zu? Mir steht ausreichend Arbeitszeit für den *Erwerb* neuer Kenntnisse zur Verfügung“

(7b) „Mir steht ausreichend Arbeitszeit zur Verfügung, um neue Kenntnisse zu *erwerben*“.

2.5 Vermeiden von hypothetischen Fragen

Um eine hypothetische Frage handelt es sich immer dann, wenn sich die Befragten zur Beantwortung der Frage in eine Situation hineinversetzen müssen, die nicht ihrer Realität entspricht, wie z.B.:

- (9) „Angenommen Sie wären Bundeskanzlerin, welches Problem in diesem Land würden Sie als erstes angehen?“
- (10) „Angenommen Sie hätten eine erwachsene Tochter, würden Sie es ihr wünschen, dass sie Karriere macht?“

Problematisch sind hypothetische Fragen vor allem deshalb, weil sie von den Befragten erfordern, sich gedanklich zunächst in eine imaginäre Situation hineinzuversetzen, diese im Kurzzeitgedächtnis zu behalten und dann die Frage aus dieser hypothetischen Situation heraus zu beantworten. Wie gut diese mentale (Doppel-)Leistung von Befragten erbracht werden kann, hängt zum einen von ihren kognitiven Fähigkeiten ab, zum anderen aber auch davon, inwiefern sie die Befragten überhaupt schon einmal mit der hypothetischen Situation auseinandergesetzt haben bzw. wie weit die hypothetische von der realen Situation entfernt ist (Porst, 2008).

Anstelle Fragen hypothetisch zu formulieren kann man die Bewertung der interessierenden Sachverhalte im Hier und Jetzt und aus der Sicht der befragten Personen abfragen. So ist es nicht nötig anzunehmen, dass man Bundeskanzlerin ist, um die Priorität eines Problems zu bewerten. Man kann hier fragen: „Welches Problem sollte von der Bundesregierung als erstes angegangen werden?“

2.6 Vermeiden von doppelten Stimuli

Fragen, die zwei oder gar mehr inhaltlich unterschiedliche Sachverhalte (Stimuli) ansprechen, sind für Befragte immer dann schwer zu beantworten, wenn diese die einzelnen Stimuli unterschiedlich bewerten. Die folgende Frage (11) enthält beispielsweise vier Stimuli:

- (11) „Inwieweit stimmen Sie der folgenden Aussage zu?
Der Staat muss dafür sorgen, dass man auch bei *Krankheit, Not, Arbeitslosigkeit* und im *Alter* ein gutes Auskommen hat.“

Befragte, die für eine staatliche Unterstützung in allen vier Bereichen sind, haben hier keine Schwierigkeiten zu antworten. Sie können einfach mit „Stimme voll und ganz zu“ oder „Stimme zu“ antworten. Anders verhält es sich bei Befragten, die zwar staatliche Unterstützung im Falle von „Krankheit“ und „Not“ befürworten, nicht jedoch bei „Arbeitslosigkeit“ und im „Alter.“ Diese Befragten können die Frage mit den vorgegebenen Antwortkategorien gar nicht beantworten (ihre Antwort wäre nämlich z.B. „Bei Krankheit und Not stimme ich zu. Bei Arbeitslosigkeit und im Alter stimme ich nicht zu.“) und falls sie es doch tun, weiß man am Ende nicht, auf welche der vier Stimuli sich ihre Antworten beziehen.

Es gibt unterschiedliche Möglichkeiten, doppelte Stimuli zu vermeiden. Eine Möglichkeit besteht darin, sich für eine Alternative zu entscheiden, und nur diese in der Frage zu belassen:

- (12) „Der Staat muss dafür sorgen, dass man auch bei *Arbeitslosigkeit* ein gutes Auskommen hat“.

Eine weitere Möglichkeit besteht darin, jeden Stimulus bzw. einzelne Stimuli bewerten zu lassen, in Form von Ratings oder forced choice Fragen (Beispiel 13; siehe auch 4.2):

- (13) „In welchem Fall soll der Staat eine finanzielle Unterstützung derart leisten, dass Menschen ein gutes Auskommen haben?“

Bei Arbeitsunfähigkeit aufgrund einer Erkrankung ja/nein

Bei Arbeitslosigkeit ja/nein

Bei unzureichender Rente/Pension im Alter ja/nein"

Schließlich kann es sein, dass die Forschenden gar nicht an den Einzelheiten interessiert sind, sondern generell die Befürwortung staatlicher Unterstützung abfragen möchten, wenn Menschen keine bzw. geringe Einkünfte haben. In diesem Fall sollte die Frage allgemein gestellt werden, wobei die einzelnen Stimuli als Beispiele dienen können:

- (14) „Der Staat muss dafür sorgen, dass Menschen mit geringen Einkünften ein gutes Auskommen haben, z.B. bei Arbeitslosigkeit, Arbeitsunfähigkeit oder im Alter.“

2.7 Vermeiden von Negationen und doppelten Verneinungen

Besonders schwierig zu verstehen und zu beantworten sind Fragen mit doppelten Verneinungen. Diese verlangen von den Befragten meist die Bewältigung einer komplizierten Aufgabe in logischem Denken (Foddy, 1993; Fowler, 2001). Doppelte Verneinungen treten z.B. immer dann auf, wenn eine negativ formulierte Aussage mittels einer Zustimmungsskala (Stimme voll und ganz zu – Stimme überhaupt nicht zu) beantwortet werden soll, z.B.:

- (15a) „Inwieweit stimmen Sie der folgenden Aussage zu?
Von ärmeren Ländern sollten *nicht weniger* Anstrengungen für den Umweltschutz erwartet werden als von reichen Ländern.“

Um hier eine positive Meinung zum Thema der Aussage auszudrücken („Von ärmeren Ländern sollten weniger Anstrengungen für den Umweltschutz erwartet werden“), müssen die Befragten der negativ formulierten Aussage *nicht* zustimmen. Dafür müssen die Befragten mental zunächst die negative Aussage in eine positive Aussage „übersetzen“ und dann entscheiden, ob sie der Aussage zustimmen oder nicht zustimmen (Akiyama, Brewer, & Shoben, 1979). Positiv formulierte Aussagen sind einfacher zu verstehen und zu beantworten, da hier der „Übersetzungsschritt“ wegfällt:

- (15b) „Inwieweit stimmen Sie der folgenden Aussage zu?
Von ärmeren Ländern sollten weniger Anstrengungen für den Umweltschutz erwartet werden als von reichen Ländern.“

Negationen – Wörter wie „nicht“ oder „keine“ – sind grundsätzlich in Fragen zu vermeiden, weil sie leicht übersehen werden können (Dillman et al., 2009).

2.8 Vermeiden von (impliziten) Unterstellungen

Unterstellungen sind dann vorhanden, wenn der Fragetext eine Situation als gegeben erachtet, die bei den Befragten nicht unbedingt erfüllt sein muss. Solche Unterstellungen sollten in Survey-Fragen vermieden werden, da Befragte, welche die Unterstellung nicht teilen bzw. dieser nicht zustimmen, die Frage im Grunde nicht richtig beantworten können. Hier ein Beispiel für eine Unterstellung (vgl. Fillmore, 1999):

- (16) „Das Familienleben leidet oft, weil Männer sich zu sehr auf ihre Arbeit konzentrieren.“

Diese Aussage unterstellt, dass Männer sich zu sehr auf ihre Arbeit konzentrieren und die Befragten sollen angeben, inwiefern (bzw. wie häufig) dies ihrer Meinung nach ein Grund dafür ist, dass das Familienleben leidet. Befragte, die nicht der Meinung sind, dass Männer sich zu sehr auf ihre Arbeit konzentrieren, können diese Frage eigentlich nicht beantworten, da weder „stimme zu“ noch „stimme nicht zu“ ihre eigentliche Meinung ausdrückt.

Zu Unterstellungen kommt es auch dann, wenn die gestellte Frage auf die Befragten nicht zutrifft bzw. die Befragten in Bezug auf den Frageinhalt nicht betroffen sind (Dillman et al., 2009). Wenn man beispielsweise fragt, ob man energetische Sanierung des eigenen Hauses in Erwägung zieht, adressiert man nur die Hausbesitzer und hier auch nur diejenigen, die bisher keine energetische Sanierung vorgenommen hatten. Bei einer solchen Frage wäre angemessen, zunächst zu fragen, ob der/die Befragte Eigentümer/in des Hauses ist, dann ob eine energetische Haussanierung bereits vorgenommen wurde. Mit Hilfe von Filtern kann in einem nächsten Schritt die Frage nach der intendierten energetischen Haussanierung nur an die Eigentümer gestellt werden, deren Häuser noch nicht saniert sind. Bei der Nutzung der Filterfragen sollte jedoch der Modus der Datenerhebung beachtet werden. Bei einer Computerunterstützung sind Filterführungen einfach umzusetzen und bereiten keine Schwierigkeiten für die Befragten. Bei schriftlich-postalischen Befragungen hingegen sollten zahlreiche und komplexe Filter vermieden werden, was generell die Abfragbarkeit bestimmter und zu detaillierter Informationen einschränkt (vgl. Menold, 2015, in diesem Webangebot).

2.9 Vermeiden von suggestiven Fragen

Suggestiv sind Survey-Fragen immer dann, wenn sie den Befragten eine bestimmte Antwort nahelegen bzw. diese provozieren. Hier ein Beispiel aus Porst (2008):

(17a) „Führende Wissenschaftler sind der Ansicht, dass Autoabgase das Wachstum von Kindern hemmen können. Halten Sie diese Ansicht für richtig, oder halten Sie diese Ansicht für falsch?“

Formulierungen wie „Führende Wissenschaftler“ oder „Die meisten Menschen“ führen häufig dazu, dass Befragte sich nicht trauen, den vorgegebenen Autoritäten oder der Mehrheit zu widersprechen. Daher können solche Formulierungen konforme Antworten provozieren (Porst, 2008). Suggestivfragen können auch dazu führen, dass Befragte sich vor allem über die Suggestion Gedanken machen und sich weniger damit beschäftigen, die Frage zu verstehen und angemessen zu beantworten, was wir oftmals in kognitiven Pretests beobachteten. Auch dies hätte wiederum einen negativen Effekt auf die Qualität ihrer Antworten. Daher sollte man Suggestionen in Survey-Fragen vermeiden und die Frage in Beispiel (17a) umformulieren:

(17b) „Halten Sie die Aussage, dass Autoabgase das Wachstum von Kindern hemmen können, für richtig oder für falsch?“

Ein Spezialfall einer suggestiven Frage liegt dann vor, wenn in einer Alternativfrage nur eine der beiden Antwortalternativen explizit formuliert ist (Faulbaum, Prüfer, & Rexroth, 2009), wie im Beispiel (18a):

(18a) „Finden Sie, dass in einem Betrieb alle Arbeiter in der Gewerkschaft sein sollten?“

Bei solchen „nicht ausbalancierten“ Fragen besteht die Gefahr, dass die genannte Alternative mehr Zustimmung erfährt als die nicht ausformulierte Alternative (Schuman & Presser, 1981). Um einen solchen Effekt zu vermeiden sollten bei dichotomen Fragen immer beide Antwortalternativen explizit formuliert sein:

(18b) „Finden Sie, dass in einem Betrieb alle Arbeiter in der Gewerkschaft sein sollten, oder muss man es jedem Einzelnen überlassen, ob er in der Gewerkschaft sein will oder nicht?“

Darüber hinaus besteht die Möglichkeit, abgestufte Antworten mit Hilfe von Ratingskalen abzufragen und die Aussage 18a auf Zustimmung bewerten zu lassen mit der einleitenden Frage: „Inwieweit stimmen Sie der folgenden Aussage zu?“ Zur Beantwortung wird in einem solchen Fall eine Ratingskala vorgelegt (vgl. Saris & Gallhofer, 2007; Dillman et al., 2009).

3. Wie kann man das Erinnern von relevanten Informationen und die Urteilsbildung erleichtern?

Im Idealfall haben Befragte die Antwort auf eine Frage direkt verfügbar und müssen diese nur aus dem Gedächtnis abrufen. Dieser Idealfall ist jedoch eher die Ausnahme als die Regel und kann lediglich bei sozio-demographischen Fragen wie z.B. „In welchem Monat und Jahr sind Sie geboren?“ oder „Wie viele Kinder unter 18 Jahren haben Sie?“ erwartet werden (vgl. Porst, 2008). Bei der Beantwortung der allermeisten Survey-Fragen müssen Befragte jedoch zunächst nach relevanten Informationen in ihrem Gedächtnis suchen und dann im Anschluss eine Antwort auf die Frage generieren.

Bei *Einstellungsfragen* müssen Befragte entweder eine bereits gebildete Meinung erinnern und abrufen oder – falls keine solche Meinung vorliegt – relevante Informationen erinnern, die es ermöglichen, eine Meinung unmittelbar in der Befragungssituation zu bilden. Wie leicht oder schwer dies Befragten fällt hängt vor allem davon ab, wie wichtig ihnen das Thema persönlich ist, wie stark die Einstellungen zum Thema ausgeprägt sind und wie viel Erfahrungen sie überhaupt mit dem Thema der Frage haben. Für Konsumenten von Marihuana ist die Frage „Sollte Ihrer Meinung nach der Konsum von Marihuana legalisiert werden oder nicht?“ höchstwahrscheinlich einfacher zu beantworten als für Menschen, die sich noch nie oder bisher kaum mit dem Thema der Legalisierung weicher Drogen beschäftigt haben. Als Fragebogenentwickler hat man hier (über die Empfehlungen aus dem vorherigen Abschnitt hinaus) wenig Möglichkeiten, das Erinnern von relevanten Informationen durch die Art der Frageformulierung zu erleichtern.

Anders verhält es sich mit *Verhaltensfragen* (z.B. „Wie häufig waren Sie im letzten Jahr beim Zahnarzt?“). Um diese zu beantworten, müssen Befragte relevante Ereignisse erinnern, diese gegebenenfalls korrekt datieren und die Anzahl der Ereignisse zählen oder schätzen. Wie leicht relevante Informationen erinnert werden, hängt vor allem von der Zeit ab, die seit dem Ereignis, das in einer Frage abgefragt wird, vergangen ist und von der Bedeutung des Ereignisses für den Befragten (Cannell, Marquis, & Laurent, 1977). Dies hat eine Reihe von Implikationen für die Formulierung von Survey-Fragen.

3.1 Einen passenden Zeitrahmen für die Frage wählen

Der zeitliche Rahmen einer Survey-Frage sollte vergleichsweise kurz sein, wenn:

- es sich beim Inhalt der Frage um eher alltägliche Dinge mit geringer Bedeutung handelt, z.B. „Wie häufig haben Sie in den letzten *5 Tagen* (anstatt z.B. *30 Tagen*) Fertiggerichte gegessen?“
- ein häufiges aber unregelmäßiges Verhalten erfragt wird, z.B. „Wie viele SMS haben Sie in den letzten *3 Tagen* (anstatt z.B. *4 Wochen*) versendet?“
- ein regelmäßiges Verhalten erfragt wird, z.B. „Wie oft waren Sie in den letzten *3 Monaten* (anstatt z.B. *12 Monaten*) beim Friseur?“ (Hier jedoch vorausgesetzt, dass Personen regelmäßig und nicht zu selten zu einem Friseur gehen).

Ist der zeitliche Bezug einer Frage eher kurz, so zählen die Befragten einfach die Anzahl der Ereignisse und berichten die Zahl, an die sie sich erinnern. Bei längeren zeitlichen Bezügen zählen sie typischerweise nur für eine kurze Zeitperiode und errechnen dann auf Basis dieser Periode die Zahl für die gesamte Zeitspanne. Folglich erhält man durch längere Zeitrahmen nicht wie angenommen mehr Informationen, da diese ja lediglich auf einer kürzeren Periode basieren.

Es gibt allerdings auch Situationen, in denen eher ein längerer zeitlicher Rahmen verwendet werden sollte, nämlich wenn:

- es sich bei der Frage um sehr bedeutsame und wichtige Dinge handelt, die man leicht erinnert und die sich selten ereignen, z.B. „Wie häufig sind Sie in den *letzten 10 Jahren* umgezogen?“ oder „Wie häufig waren Sie *in Ihrem Leben* bisher verheiratet?“
- ein eher seltenes und unregelmäßiges Verhalten erfragt wird, z.B. „Wie oft haben Sie in den *letzten 12 Monaten* an einer ärztlichen Vorsorgeuntersuchung teilgenommen?“

Würde man bei solchen Verhaltensfragen kürzere Zeiträume wählen, so würde man höchstwahrscheinlich eine sehr große Anzahl an „0“-Antworten erhalten.

3.2 Eindeutige zeitliche Bezüge verwenden

Generell sollten zeitliche Bezüge immer so eindeutig wie möglich sein. Der zeitliche Rahmen von Frage (20a) kann z.B. entweder wie in (20b) oder sogar wie in (20c) konkretisiert werden:

(20a) „Haben Sie *kürzlich* einen oder mehrere Ärzte aufgesucht?“

(20b) „Haben Sie in den *letzten vier Wochen* einen oder mehrere Ärzte aufgesucht?“

(20c) „Haben Sie *seit dem 1. April 2014* einen oder mehrere Ärzte aufgesucht?“

Die Angabe eines Datums konkretisiert zwar den zeitlichen Rahmen der Frage am deutlichsten, dies garantiert jedoch nicht, dass die Befragten sich auch leichter an die relevanten Informationen erinnern. Kalendarische Daten sind im menschlichen Gedächtnis nur schlecht repräsentiert und meist nicht als Teil von Ereignissen abgespeichert. Möglicherweise kann man sich z.B. noch gut an die Situation erinnern, als man sein vorletztes Fahrrad gekauft hat, kann das Datum dieses Ereignisses aber nicht mehr erinnern (vgl. Porst, 2008).

3.3 Aufteilung von einer komplexen Frage in mehrere Einzelfragen

Bei komplexen Fragen kann die Aufteilung von einer in mehrere Fragen das Erinnern relevanter Informationen und die Urteilsbildung erleichtern. Die Frage „Wie viele verschiedene Ärzte haben Sie in den letzten 12 Monaten aufgesucht?“ kann beispielsweise bei einer relativ großen Anzahl an Arztbesuchen schwierig zu beantworten sein und es ist wahrscheinlich, dass der ein oder andere Arztbesuch nicht erinnert wird. Daher bietet es sich an, die Frage in mehrere Fragen zu unterteilen und diese nacheinander zu fragen (Fowler & Cosenza, 2008):

(21a) „Haben Sie in den letzten 12 Monaten einen oder mehrere *Hausärzte/Allgemeinmediziner* aufgesucht? (Falls ja: Wie viele?)“

(21b) „Haben Sie in den letzten 12 Monaten einen oder mehrere *Fachärzte* aufgesucht? (Falls ja: Wie viele?)“

(21c) „Haben Sie in den letzten 12 Monaten einen oder mehrere *Heilpraktiker* aufgesucht? (Falls ja: Wie viele?)“

(21d) „Wie viele *andere Ärzte*, die Sie noch nicht erwähnt haben, haben Sie letzten 12 Monaten aufgesucht?“

Ein solches Vorgehen erleichtert den Befragten das Erinnern der relevanten Informationen und die Urteilsbildung. Zum einen müssen sie sich jeweils nur auf einen Teilbereich der relevanten Informationen beschränken (in diesem Beispiel auf einen bestimmten Teil von Ärzten) und nicht an verschiedene Bereiche gleichzeitig denken. Zum anderen haben die Befragten so auch mehr Zeit, sich über die einzelnen Fragen Gedanken zu machen. Dadurch wird es ihnen leichter fallen, die Informationen zu erinnern. Zu guter Letzt erfüllen die vorherigen Fragen in einem solchen gestuften

Vorgehen eine gewisse „Priming“-Funktion: Dadurch, dass ähnliche Informationen (z.B. Fachärzte) zuvor bereits erinnert wurden, fällt es den Befragten leichter, die gesuchte Informationen (z.B. Heilpraktiker) zu erinnern.

4. Wie kann man das Einpassen der „intern“ ermittelten Antwort in das vorgegebene Antwortformat erleichtern?

Die letzte Aufgabe des Befragten bei der Beantwortung einer Survey-Frage besteht darin, die „intern“ ermittelte Antwort auf die Frage in der vom Fragebogen (bzw. vom Interviewer) vorgegebenen Form zu kommunizieren. Die Wahl und das Design des Antwortformats kann dabei einen starken Effekt auf die Beantwortung der Frage haben (Couper, Tourangeau, Conrad, & Crawford, 2004; Tourangeau, Couper, & Conrad, 2004, 2007).

Im Allgemeinen kann man drei große Gruppen von Antwortformaten unterscheiden:

1. Offene Fragen
Diese Fragen enthalten lediglich den Fragetext und eventuell eine Ausfüllanweisung, jedoch keine Antwortkategorien. Befragte beantworten offene Fragen also in ihren eigenen Worten. (siehe SDM Survey Guidelines Kapitel „Offene Fragen“ (Züll, 2015))
2. Geschlossene Fragen mit Ratingskalen
Diese Fragen enthalten eine vordefinierte Anzahl von Antwortmöglichkeiten, welche in einer inhärenten Logik angeordnet sind und sich auf eine bestimmte Dimension beziehen, z.B. den Grad der Stärke bzw. die Intensität („sehr stark – stark – mittel – wenig – überhaupt nicht“). Befragte sollen mittels der mehrstufigen Ratingskala zum Ausdruck bringen, wo sie sich in Bezug auf ein bestimmtes Merkmal (z.B. Interesse an Politik) auf der Antwortdimension verorten (siehe SDM Survey Guidelines Kapitel „Gestaltung von Ratingskalen in Fragebögen (Menold & Bogner, 2015)).
3. Geschlossene Fragen ohne Ratingskalen
Diese Fragen enthalten ebenfalls eine vordefinierte Anzahl von Antwortmöglichkeiten, allerdings nicht in Form einer Ratingskala, sondern lediglich in Form einer Liste. Die Anordnung der Antwortmöglichkeiten kann entweder einer inhärenten Logik folgen (wie z.B. bei der Frage nach dem höchsten allgemeinbildenden Schulabschluss, bei der die Antwortmöglichkeiten vom geringsten bis zum höchsten möglichen Schulabschluss angeordnet sind) oder dem Fragebogenentwickler überlassen sein (wie z.B. bei der „Sonntagsfrage“ nach der Parteipräferenz, bei der die Antwortkategorien entweder „CDU, SPD, Grüne, FDP...“ oder aber auch „SPD, CDU, FDP, Grüne...“ angeordnet sein können).

Da offene Fragen und geschlossene Fragen mit Ratingskalen in anderen Kapiteln detailliert behandelt werden, sollen hier nur allgemeine Hinweise zur Formulierung der Antwortkategorien in geschlossenen Fragen gegeben werden, wobei jeweils auf die Aspekte der Ratingskalen wie z. B. Anzahl der Kategorien oder Wahl der Mittelkategorie nicht eingegangen wird.

4.1 Die Antwortkategorien sollten zur Frage passen

Hin und wieder kommt es vor, dass die Antwortoptionen einer Frage nicht richtig zur eigentlichen Frage passen, wie in folgendem Beispiel:

- (22) „Bitte geben Sie für jede der folgenden Organisationen an, ob Sie in den letzten 12 Monaten als aktives Mitglied in diesen Organisationen tätig waren.“

- a) Sportverein
- b) Gewerkschaft
- c) Politische Partei

Antwortkategorien: Oft – Manchmal – Selten – Nie“

Im Fragetext geht es darum, „ob“ man in den letzten 12 Monaten aktiv in einer der Organisationen tätig war. Die Frage ist also so gestellt, dass man sie eigentlich nur mit „Ja“ oder „Nein“ beantworten kann. Die Antwortkategorien verlangen jedoch die Angabe einer Häufigkeit. Dieser Widerspruch kann bei Befragten zu Irritationen führen, daher sollte entweder der Fragetext den Antwortkategorien angepasst werden („Wie häufig waren Sie...“) oder die Antwortkategorien dem Fragetext angepasst werden („Ja“, „Nein“).

4.2 Die Antwortkategorien sollten erschöpfend und überschneidungsfrei sein

Bei geschlossenen Antworten müssen die vordefinierten Antwortoptionen das gesamte Spektrum an möglichen Antworten abdecken. Falls inhaltlich-logische Antwortoptionen fehlen, werden einige Befragte ansonsten die Frage nicht beantworten oder willkürlich eine der Antwortkategorien auswählen (Faulbaum et al., 2009). In der folgenden Frage (23) sind nur wenige Bereiche aufgelistet, die von den Befragungsteilnehmenden als wichtige Aufgaben der Bundesregierung aufgefasst werden könnten.

- (23) „Bei der folgenden Frage geht es um die kürzlich neu gebildete Bundesregierung in Deutschland. Welchen politischen Aufgaben sollte sich die neue Bundesregierung Ihrer Meinung nach in den nächsten vier Jahren verstärkt widmen? Bitte geben Sie alles Zutreffende an!

Entwicklungshilfen für arme Länder
 Subventionen in der Landwirtschaft
 Verteidigung und hier speziell der Abzug der Streitkräfte aus Afghanistan
 Finanzkrise in der Europäischen Union“

Bei dieser Frage sollte die Liste um weitere bedeutende Aspekte ergänzt werden, wie z.B. Frauenförderung, Ausstieg aus der Atomkraft, Finanzielle Unterstützung der Familien mit Kindern, Ausbau der Krippenplätze. Es ist außerdem kaum möglich, alle möglichen Alternativen in diesem Fall aufzulisten, so dass die Liste um eine „Sonstiges“-Kategorie ergänzt werden sollte.

Fragen, die eine bestimmte Anzahl von Alternativen auflisten und entweder die Auswahl nur einer Alternative (sog. Einfachnennungsfragen) oder mehrerer Alternativen (sog. Mehrfachnennungsfragen) zulassen, sind in Umfragen sehr häufig. Bei solchen Antwortlisten sind Reihenfolge-Effekte der Darbietung möglich. So können bei einer visuellen Präsentation die ersten Alternativen in der Liste häufiger als die nachfolgenden Alternativen gewählt werden (primacy-effect). Bei der akustischen Präsentation hingegen werden die letztgenannten Alternativen häufiger gewählt (recency-effect). Um solche Reihenfolge-Effekte zu vermeiden, sollte jede Alternative mit einer „ja/nein“ Vorgabe (sog. forced choice, siehe Smyth et al., 2006) beantwortet werden. Dieses Design ist z. B. auch in der obigen Frage (22) realisiert.

Außerdem muss bei geschlossenen Fragen sichergestellt werden, dass die Antwortkategorien überschneidungsfrei sind, so dass Befragte ihre Antworten zweifelsfrei einer Kategorie zuordnen können. In Frage (24) können z.B. Befragte, die morgens um 6:30 Uhr aufstehen, sowohl die zweite als auch die dritte Antwortkategorie auswählen:

- (24) „Um welche Uhrzeit stehen Sie üblicherweise an einem Werktag auf?

- Vor 6:00 Uhr (vormittags)
- Zwischen 6:00 Uhr und 6:30 Uhr (vormittags)
- Zwischen 6:30 Uhr und 7:00 Uhr (vormittags)
- Zwischen 7:00 Uhr und 7:30 Uhr (vormittags)
- Zwischen 7:30 Uhr und 8:00 (vormittags)
- Nach 8:00 (vormittags)"

Die Antwortoptionen sollten daher umformuliert werden, so dass sie überschneidungsfrei sind:

(25) „Um welche Uhrzeit stehen Sie üblicherweise an einem Werktag auf?"

- Vor 6:00 Uhr (vormittags)
- Zwischen 6:00 Uhr und 6:29 Uhr (vormittags)
- Zwischen 6:30 Uhr und 6:59 Uhr (vormittags)
- ...

4.3 Die Antwortkategorien sollten logisch/systematisch angeordnet sein

Um Befragten das Verständnis des Antwortspektrums zu erleichtern, sollten Antwortkategorien, denen eine gewisse Abstufungslogik unterliegt, entsprechend dieser Logik angeordnet sein. Die Antwortreihenfolge in der folgenden Frage (26) ist beispielsweise inhaltlich nicht logisch:

(26) „Sind Ihre Beschwerden durch die Einnahme der Medikamente..."

- ganz weg,
- etwas besser,
- deutlich besser, oder
- unverändert?"

Die Antwortkategorien „etwas besser" und „deutlich besser" müssten hier vertauscht werden, um eine logische Abstufungsreihenfolge herzustellen.

4.4 Die Antwortkategorien sollten keine Regelmäßigkeit unterstellen

Manche Verhaltensfragen oder Faktfragen unterstellen, dass das erfragte Verhalten oder die erfragten Ereignisse einer gewissen Regelmäßigkeit unterliegen. Dies muss für Befragte jedoch nicht immer zutreffen, was die Beantwortung der Frage für diese Befragten dann erschwert (Fowler & Cosenza, 2008). Hier ein Beispiel:

(27a) „Wie häufig heben Sie Geld an einem Geldautomaten ab?"

- Gar nicht
- Seltener als einmal pro Monat
- Einmal pro Monat
- Alle zwei Wochen
- Jede Woche"

In der Frage (27a) wird beispielsweise unterstellt, dass Befragte in regelmäßigen Abständen am Geldautomaten Geld abheben. Dies muss in der Realität aber nicht für alle Befragte zutreffen, sie könnten ebenso auch nur auf Bedarfsbasis (und in unregelmäßigen Abständen) Geld am Geldautomaten abheben. Daher sollte man hier besser nach der Anzahl der Gänge zum Geldautomaten innerhalb eines konkreten Zeitrahmens fragen. Frage (27b) unterstellt beispielsweise kein regelmäßiges Verhalten:

(27b) „Wie häufig haben Sie im letzten Monat Geld an einem Geldautomaten abgehoben?"

- Gar nicht

- Einmal
- Zwei- bis dreimal
- Viermal und häufiger"

5. Zusammenfassung

Aus der obigen Darstellung wird deutlich, dass bei der Formulierung von Survey-Fragen eine Vielzahl methodischer Aspekte zu berücksichtigen ist. Die Berücksichtigung dieser Aspekte ist eine Grundvoraussetzung dafür, dass Survey-Fragen von den Befragten in der intendierten Weise interpretiert werden und möglichst leicht zu verstehen und zu beantworten sind. Oft besteht jedoch auch dann Unsicherheit in Bezug auf die Verständlichkeit und die Beantwortbarkeit von Fragen, wenn scheinbar alle besprochenen formalen Regeln berücksichtigt sind. Mehr Sicherheit darüber, ob eine Frage im Fragebogen wie intendiert funktioniert und einfach zu beantworten ist, erhält man zumeist erst durch die Durchführung eines kognitiven Pretests (vgl. Lenzner, Neuert und Otto, 2015 in diesem Webangebot). Selbst mit Hilfe kognitiver Pretests kann man jedoch nicht zweifelsfrei sicherstellen, dass eine Frage zuverlässig (Reliabilität) die intendierten Inhalte (Validität) erfasst. Zur Untersuchung der Reliabilität und Validität von Survey-Fragen sind weitere theoretische und empirische Schritte notwendig. Hierzu wird an dieser Stelle auf den Bereich „Qualitätssicherung“ dieses Webangebots und auf die Qualitätsstandards des RatSWD (Rammstedt et al., 2014) verwiesen.

Literaturverzeichnis

- Akiyama, M. M., Brewer, W. F., & Shoben, E. J. (1979). The yes-no question answering system and statement verification. *Journal of Verbal Memory and Verbal Behavior*, 18, 365-380.
- Cannell, C. F., Marquis, K. H., & Laurent, A. (1977). A summary of studies of interviewing methodology: 1959-1970. *Vital and Health Statistics*, 69, i-78.
- Cannell, C., Miller, P., & Oksenberg, L. (1981). Research on interviewing techniques. In S. Leinhardt (Ed.), *Sociological methodology 1981* (pp. 389-437). San Francisco: Jossey-Bass.
- Couper, M. P., Tourangeau, R., Conrad, F. G., & Crawford, S. D. (2004). What they see is what we get. *Social Science Computer Review*, 22, 111-127.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys. The tailored design method*. Wiley: New Jersey.
- Faulbaum, F., Prüfer, P., & Rexroth, M. (2009). *Was ist eine gute Frage? Die systematische Evaluation der Fragenqualität*. Wiesbaden: VS Verlag.
- Fillmore, C. J. (1999). A linguistic look at survey research. In M. G. Sirken, D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur, & R. Tourangeau (Eds.), *Cognition and survey research* (pp. 183-198). New York: Wiley.
- Foddy, W. (1993). *Constructing questions for interviews and questionnaires: theory and practice in social research*. Cambridge: Cambridge University Press.
- Fowler, F. J. (1992). How unclear terms affect survey data. *Public Opinion Quarterly*, 56, 218-231.
- Fowler, F. J. (2001). Why it is easy to write bad questions. *ZUMA-Nachrichten*, 48, 49-66.

- Fowler, F. J., & Cosenza, C. (2008). Writing effective questions. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 136-160). New York: Lawrence Erlbaum.
- Graesser, A. C., Cai, Z., Louwse, M. M., & Daniel, F. (2006). Question understanding aid (QUAID). A web facility that tests question comprehensibility. *Public Opinion Quarterly*, *70*, 3-22.
- Lenzner, T. (2011). *A psycholinguistic look at survey question design and response quality*. University of Mannheim: MADOC.
- Lenzner, T. (2012). Effects of survey question comprehensibility on response quality. *Field Methods*, *24*, 409-428.
- Lenzner, T., Kaczmirek, L., & Galesic, M. (2011). Seeing through the eyes of the respondent: An eye-tracking study on survey question comprehension. *International Journal of Public Opinion Research*, *23*, 361-373.
- Lenzner, T., Kaczmirek, L., & Lenzner, A. (2010). Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology*, *24*, 1003-1020.
- Lenzner, T., Neuert, C. & Otto, W. (2015). Kognitives Pretesting. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (*SDM Survey Guidelines*). DOI: 10.15465/sdm-sg_010.
- Menold, Natalja (2015). Schriftlich-postalische Befragung. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (*SDM Survey Guidelines*). DOI: 10.15465/sdm-sg_018
- Menold, Natalja und Bogner, Kathrin (2015). Gestaltung von Ratingskalen in Fragebögen. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (*SDM Survey Guidelines*). DOI: 10.15465/sdm-sg_015
- Porst, R. (2008). *Fragebogen. Ein Arbeitsbuch*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Rammstedt, B., Beierlein, C., Brähler, E., Eid, M., Hartig, J., Kersting, M., Liebig, S., Lukas, J., Mayer, A.-K., Menold, N., Schupp, J. & Weichselgartner, E. (2014). Qualitätsstandards zur Entwicklung, Anwendung und Bewertung von Messinstrumenten in der sozialwissenschaftlichen Umfrageforschung. In: *RatSWD Working Papers Series*, 230, S. 1-37.
- Saris, W. E., & Gallhofer, I. N. (2007). *Design, evaluation, and analysis of questionnaires for survey research*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. New York: Academic Press.
- Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006). Comparing Check-All and Forced Choice Question Formats in Web Surveys: The Role of satisficing, depth of processing, and acquiescence in explaining differences. *Public Opinion Quarterly*, *70*, 66-77.
- Tourangeau, R. (1984). Cognitive science and survey methods. In T. B. Jabine, M. L. Straf, J. M. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 73-100). Washington, DC: National Academy Press.
- Tourangeau, R., Couper, M. P., & Conrad, F. G. (2004). Spacing, position, and order: Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly*, *68*, 368-393.
- Tourangeau, R., Couper, M. P., & Conrad, F. G. (2007). Color, labels, and interpretive heuristics for response scales. *Public Opinion Quarterly*, *71*, 91-112.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

Züll, Cornelia (2015). Offene Fragen. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (SDM Survey Guidelines). DOI: 10.15465/sdm-sg_002