

To force or not to force. That is the question!

Die Auswirkungen des Einsatzes von Forced Answering-Fragen auf die Qualität der Befragungsergebnisse

Alexandra Mergener, Philipp Sischka und Jean Philippe Décieux

Beitrag zur Postersession

Die Methode der Onlinebefragung hat sich innerhalb des letzten Jahrzehnts als Befragungsroutine etabliert. Sie ist zumeist mit niedrigen Kosten verbunden und ermöglicht es innerhalb kürzester Zeit hohe Fallzahlen zu erzielen sowie »fundierte« Ergebnisse zu generieren. Dies führte einerseits zu einer »Demokratisierung der Umfrageforschung« (Batinic 2003: 6), denn mithilfe der Onlinebefragung ist es nahezu jedem möglich ein Befragungsprojekt durchzuführen. Andererseits resultiert daraus aber auch, dass viele Befragungen von Laien umgesetzt werden, dadurch häufig eine schlechte Qualität aufweisen und zahlreiche Operationalisierungsfehler enthalten (Jacob et al. 2013; Schnell et al. 2011).

Ein Beispiel hierfür ist die Verwendung der Forced Answering-Option (FA)¹, welche die Möglichkeit bezeichnet den Respondenten einer Umfrage zur Beantwortung einer Frage zu verpflichten. Bei den meisten Programmpaketen zur Durchführung eines Onlinesurveys ist dies auf einfache Art und Weise zu realisieren. Diese Praxis kommt zuweilen sehr häufig auch zur Anwendung, ohne dass sich die Fragesteller über eventuelle Folgen ihrer Wahl bewusst sind. In den Handbüchern einiger Programme (Questback 2013) und auch in einigen wissenschaftlichen Publikationen (Albaum et al. 2010; Roster et al. 2014) wird diese Option als eine Lösung angesehen, die den Item Nonresponse verringert.

Aus unserer Perspektive gibt es jedoch zahlreiche Bedenken bei der Implementierung dieses methodischen Features. Diese Bedenken speisen sich aus der Überlegung, dass der Befragte plausible Gründe haben kann, um nicht zu antworten (bspw. versteht der Befragte die Frage nicht; fehlt die entsprechende Antwortkategorie, will der Befragte die Frage aus persönlichen Gründen bewusst nicht beantworten, usw.). Verpflichtet man den Befragten unter diesen Umständen eine Antwort zu geben, könnte es möglicherweise dazu führen, dass die Befragung abgebrochen wird oder der Befragte eine willkürliche/inhaltsunabhängige Antwort gibt.

¹ Häufig wird auch von Forced Response (FR) gesprochen.

Wir sind damit argumentativ auf einer Linie mit Empfehlungen von zahlreichen methodologischen Standardwerken (Dillman et al. 2014; Jacob et al. 2013; Kaczmirek 2005; Peytchev, Crawford 2005; Schnell et al. 2011). Diese warnen zwar vor der Nutzung dieser Option auf Basis von praktischem Erfahrungswissen, liefern jedoch wenig empirische Evidenz für ihre Empfehlung.

Die Forced Answering-Option:

Es gibt eine Fülle an methodologischen Studien, die sich mit den Auswirkungen von unterschiedlichen Designelementen beziehungsweise -features bei Onlineumfragen und ihren Konsequenzen für den Befragungsverlauf, die Datenqualität, die Abbruchquoten etc. beschäftigt haben (Fan, Yan 2010; Galesic, Bosnjak 2009; Ganassali 2008; Marcus et al. 2007; Sakshaug et al. 2010; Vicente, Reis 2010).

Eine Option, der hingegen bisher wenig Beachtung geschenkt wurde, ist die Möglichkeit, den Befragten in einer Onlineumfrage den Fragebogen erst dann weiter bearbeiten zu lassen, nachdem er die Fragen der aktuellen Seite beantwortet hat. Dies mag zum Teil damit zusammenhängen, dass die Option sehr spezifisch für Onlinebefragungen ist. In Offline-Befragungsarten gibt es wenige bis keine wirkliche Entsprechung: Bei postalischen Befragungen ist die Implementierung einer solchen Option faktisch nicht realisierbar und auch im Rahmen einer persönlichen Interviewsituation (telefonisch oder Face-to-Face) wäre dies – so wie in Onlinebefragungen umgesetzt – gleichbedeutend mit einer Situation, in der der Interviewer das Interview nicht weiterführt, wenn der Respondent nicht auf eine gestellte Frage antworten möchte.²

Sucht man nach ähnlichen Elementen innerhalb einer Befragung, so findet man beispielsweise Gemeinsamkeiten zur Forced Choice Frageoption³, die den Befragten dazu zwingt, sich aus zwei oder mehreren Antwortalternativen für eine einzige zu entscheiden. Der Unterschied ist hier jedoch, dass Forced Choice ein spezifisches Antwortformat darstellt und die Forced Answering-Option lediglich ein Antwortfeature im Rahmen einer Befragung ist, welche komplett unabhängig vom Antwortformat der Frage ist. Die Implementierung von Pflichtfragen ist bei allen Antwortformaten möglich, beginnend bei Forced Choice Fragen, Multiple Choice Fragen sowie bei offenen Fragen (Stieger et al. 2007).

Das zentrale Motiv für die Implementierung der Forced Answering-Option ist in den meisten Fällen die Idee, einen möglichst vollständigen Datensatz ohne Item-Nonresponse zu generieren: »When used for all questions, it virtually eliminates item omissions, or item nonresponse error.« (Roster et al. 2014: 91)

² Preisendörfer 2008 implementierte in mündlichen Interviews zur Untersuchung der Erhebung heikler Fragen mittels Randomised Response Technik die Variante »forced response« in Form eines Kartenspiels. Dabei zwangen Karten mit Ja oder Nein die Befragten zu genau diesen unter Umständen nicht der Realität entsprechenden Antworten, während andere Karten mit dem Aufdruck »Wahrheit« die Befragten dazu aufforderten eine wahre Antwort zu geben. Diese Forced Response Option ist jedoch nicht direkt mit derjenigen zu vergleichen, die in diesem Beitrag untersucht wird.

³ Zuweilen werden diese häufig auch missverständlicher Weise als Synonym gebraucht.

Die Entscheidung der Nutzung dieser Option sollte jedoch mit den Kosten einer Implementierung ausbalanciert werden. Einige Forscher äußern zahlreiche Bedenken und Gegenstimmen für diese Vorgehensweise, ohne sich jedoch auf fundierte empirische Untersuchungen zu beziehen (Dillman et al. 2014; Jacob et al. 2013; Kaczmirek 2005; Peytchev, Crawford 2005; Schnell et al. 2011). Die tatsächlichen Konsequenzen der Implementierung der Forced Answering-Option werden daher in diesem Beitrag systematisch analysiert.

Bisheriger Forschungsstand

Es gibt bisher sehr wenige Studien, die den Einfluss von Forced Answering auf verschiedene Qualitätsparameter einer Onlineumfrage evaluieren. Diese fokussieren beispielsweise die Auswirkungen von Forced Answering auf die Abbruchquote, auf den Item-Nonresponse oder auf die Validität der Antworten. DeRouvray und Couper (2002) untersuchen den Einfluss von Warnmeldungen, die den Befragten angezeigt werden, wenn sie keine Antwort geben und stattdessen weiterklicken. Dabei kommen sie zu dem Ergebnis, dass dadurch Item-Nonresponse verringert werden kann (Derouvray, Couper 2002). O'Neil et al. (2003) untersuchen in mehreren Studien die Auswirkungen von verschiedenen Methodendesigns auf die Abbruchquote und die Stichprobenszusammensetzung. Dabei stellen sie einen minimalen Effekt auf die Abbruchquote fest, wenn auf der letzten Seite die Angaben zur Demographie erzwungen werden (O'Neil et al. 2003). Stieger et al. (2007) führten ein Online-Experiment mit Studenten durch, um den Einfluss von Forced Answering auf Unit-Nonresponse zu untersuchen. Wenn die Studenten eine Frage nicht beantworteten, bekamen Sie eine Aufforderung dies zu tun und konnten die Befragung erst weiterführen, wenn sie eine Antwort gaben. Dabei verglichen sie in ihrer Studie diejenigen Studenten, die mindestens einmal eine Fehlermeldung bekommen haben, mit denjenigen, die immer alle Antworten gegeben haben und daher niemals eine Fehlermeldung gesehen haben. Sie stellen deutlich höhere Abbruchquoten in der Forced Answering-Gruppe (64,1%) als in der Nicht Forced Answering-Gruppe (NFA) (18,6%) fest. Sie kommen weiterhin zu dem Ergebnis, dass sich die Forced Answering-Option als sogenannte »hurdle technique« eignet. Solche können zu Beginn einer Umfrage genutzt werden, um wenig motivierte Respondenten zum Abbruch der Befragung zu bewegen und dadurch auszusortieren (Stieger et al. 2007). Ganassali (2008) untersucht in seiner Studie den Einfluss bestimmter Designeffekte auf verschiedene Parameter der Datenqualität (u.a. Abbruchquote). Dabei stellt er fest, dass eingebaute Interaktionsfeatures (Warnmeldungen und Hinweise bei nicht beantworteten der Frage) den Item-Nonresponse verringern (Ganassali 2008). Weitere Studien wurden von Albaum und seinem Forscherteam (Albaum et al. 2014; Albaum et al. 2010; Albaum et al. 2011; Roster et al. 2014) durchgeführt. Albaum et al. (2010) untersuchen die Auswirkungen von FA auf Item- und Unit-Nonresponse, wobei sie in ihrem Design folgende Faktoren variieren: FA oder NFA, »keine Angabe«-Option (»Prefer not to answer«, PNA) verfügbar beziehungsweise nicht verfügbar, Item-Batterie Reihenfolge: 8- oder 15-Item-Block zuerst⁴. Sie stellen weder Haupteffekte noch Interak-

⁴ Die letzte Versuchsanordnung führen sie durch, da sie die Hypothese aufstellen, dass Befragte, die zunächst mit einem längeren Fragenblock konfrontiert werden, eher abbrechen werden.

tionseffekte der verschiedenen Bedingungen auf Unit-Nonresponse fest. Allerdings kommen sie auf Haupteffekte sowohl für FA als auch für die Verfügbarkeit einer ›keine Angabe‹-Option auf Item-Nonresponse. Sowohl die NFA-Option als auch die Verfügbarkeit einer ›keine Angabe‹-Option erhöhen den Item-Nonresponse: »To summarize, we find no evidence to suggest that use of forced answering, with or without PNA, decreases completion rates to Web surveys.« (Albaum et al. 2010: .292). Den Einfluss auf das Antwortverhalten versuchen sie dadurch zu ermitteln, indem sie mögliche Unterschiede in den Summenscores der Skalen untersuchen. Dabei kommen sie auf keine signifikanten Unterschiede zwischen den Versionen. Bezüglich der ›keine Angabe‹-Option kommen sie zu folgender Schlussfolgerung: »On the other hand, we do find differences in mean item omissions based on whether or not PNA is included as a response option. It appears that use of PNA along with forced answering can erode the advantages of forced answering in terms of generating a higher quantity of substantive responses to critical questions.« (Albaum et al. 2010: 291). Albaum et al. (2011) berichten über dieselbe Studie wie Albaum et al. (2010) und kommen dementsprechend zu gleichen Ergebnissen. Roster et al. (2014) ergänzen diese Studien und prüfen, ob die erzielten Ergebnisse auch bei einem sensitiven Thema auftreten. Dazu nutzen sie ein Design mit den folgenden Manipulationen: FA beziehungsweise NFA, keine Angabe-Option verfügbar oder nicht verfügbar, hoch beziehungsweise gering sensitives Thema⁵. Dieses Design nutzen sie in zwei verschiedenen Kulturkreisen (U.S. und Hong Kong). Dabei stellen sie keinen Einfluss von FA auf die Abbruchquote fest. Tendenziell ist der Unit-Nonresponse in der FA-Bedingung sogar geringer. Der Item-Nonresponse ist in der FA-Bedingung dagegen signifikant geringer, als in der NFA-Bedingung. Auch die ›keine Angabe‹-Option hat keinen Einfluss auf die Abbruchquote und den Item-Nonresponse. Daher kommen sie zu folgendem Schluss: »The findings generally support prior research concerning the lack of effects of FA, regardless of whether a PNA response option is employed.« (Roster et al. 2014: 91) Betrachtet man die Studien von Albaum et al. (Albaum et al. 2014; Albaum et al. 2010; Albaum et al. 2011; Roster et al. 2014) genauer, so fällt für die Interpretation der Ergebnisse ein nicht unerhebliches Detail auf. Die Ergebnisse seiner Studien basieren alle auf Online-Access-Panel Populationen. Die spezifische Befragungssituation in einem Online-Access-Panel ist jedoch nur mit Einschränkungen übertragbar auf die Situation einer normalen Umfrage (Göritz 2007). Dies könnte die fundamental unterschiedlichen Ergebnisse bezüglich Dropouts im Vergleich zu anderen Studien, wie zum Beispiel von Stieger, Reips, Voracek (2007) erklären.

Der kurze Überblick über den Stand der Forschung in diesem Bereich zeigt, dass es bisher keine Einigkeit bezüglich des Einflusses der Implementierung der Forced Answering-Option gibt. Die Resultate variieren zwischen positiven und negativen Einflüssen auf die Qualität der Daten. Hier besteht also ein enormes Forschungsdesiderat, welches sowohl für die Wissenschaft als auch die Umfragepraxis von großer Relevanz ist. Daher benötigt es vor allem Studien, die die Folgen von Forced Answering unter realistischen Bedingungen testen. An dieser Stelle setzt

⁵ Das Thema mit geringer Sensitivität behandelte Fragen zum Markenimage von Coca-Cola, das sensitive Thema kreiste um Fragen zu Familieneinkommen und persönlichen Finanzen (Albaum et al. 2014: 151). Den Grad der Sensitivität stellten sie mittels Manipulationscheck mit folgender Frage fest: »To what extent did you find the items listed in question 2 above [the topic manipulation question] to be sensitive (i.e. intrusive or threatening) to you personally?« (Albaum et al. 2014: 154).

dieser Beitrag an. Das zentrale Ziel der Studie ist es die Konsequenzen der Implementierung der Forced Answering-Option zu testen.

Forschungsfrage und Hypothesen

Im Rahmen des folgenden Experiments wollen wir die genauen Folgen der Implementierung einer Forced Answering-Option empirisch abbilden. Unsere zentrale These ist, dass der Zwang des Befragten eine Antwort zu geben, dazu führen kann, dass dieser die Befragung abbrechen oder eine willkürliche/ inhaltsunabhängige Antwort gibt.

Hierzu wurden den Befragten verschiedene kognitiv und im Antwortprozess mehr oder weniger aufwändige Frageformen (bspw. Itembatterien, semantisches Differential, offene Fragen etc.) mit unterschiedlicher Sensitivität präsentiert. Weiterhin soll der Einfluss einer neutralen Antwortkategorie »keine Angabe« untersucht werden. Über diese zusätzliche Antwortoption werden die Befragten von der »Pflicht«, auf die Fragen zu antworten entbunden. Die Folgen sollen über die Analyse von Abbruchquoten und Antwortverhalten abgebildet werden.

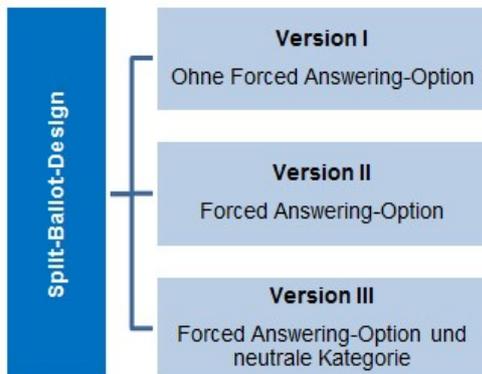
Wir wollen diesbezüglich drei Hypothesen überprüfen:

1. Forced Answering führt zu erhöhtem Unit-Nonresponse.
2. Eine neutrale Antwortkategorie kann den negativen Effekt der Forced Answering-Option auf den Unit-Nonresponse einschränken.
3. Forced Answering führt zu weniger validen Antworten (Lügen oder Zufallsantworten).

Design

Um dies zu analysieren, haben wir ein Split-Ballot-Feld-Experiment durchgeführt. Split-Ballot-Experimente bieten die Möglichkeit Designeffekte im Gegensatz zu Laborexperimenten unter möglichst realen Bedingungen (extern valide) zu testen (Décieux 2012: 15). Im Rahmen unseres Experimentes findet eine zufällige Zuordnung der Befragungspersonen in Subgruppen, die unterschiedliche Versionen der Befragung durchlaufen, statt. Respondenten, die die Befragung in Version 1 durchlaufen, nehmen an der Befragung unter ganz normalen Umständen teil; es besteht bei keiner Frage ein Antwortzwang. In Version 2 können die Respondenten die nächste Seite des Fragebogens erst sehen, wenn sie bei allen Fragen einer Seite mindestens eine der möglichen Antwortkategorien ausgewählt haben. Auch in Version 3 des Experimentes sind die Befragten gezwungen eine Antwortkategorie auszuwählen, allerdings ist diese Version um die Antwortoption »keine Angabe« (»Prefer not to answer«, PNA) erweitert, so dass die Respondenten — zumindest theoretisch — vom Zwang Informationen freizugeben befreit sind. Daraus ergibt sich das in Abb. 1 dargestellte Forschungsdesign.

Abb. 1: Forschungsdesign



Um einen möglichst hohen Rücklauf zu erreichen, haben wir ein inhaltliches Thema gewählt, das allgemein und bekannt ist, sodass möglichst viele Personen als Respondenten infrage kommen. Das Leitthema der Befragung war «Kriminalität und ihre Wahrnehmung in der Bevölkerung». Hier nutzten wir das Steuerdelikt von Uli Hoeneß, das zu dieser Zeit ein großes Thema in den Medien war als Aufhänger, um ein möglichst großes Interesse zu wecken und dadurch eine entsprechend hohe Teilnahmequote zu erreichen. Weiterhin wurden Incentives in einem Gesamtwert von 350 € verlost. Die Befragung beinhaltete neben Fragen zu Uli Hoeneß (Bewertung der Person, Bewertung der Straftat, etc.) und zur Medienpräsenz dieses Falles auch einige sehr sensible Items zum eigenen delinquenten Verhalten sowie zu eigenen Opfererfahrungen.

Die Befragung startete am 12.06.2014 (unmittelbar nach der Verurteilung von Ulli Hoeneß) und dauerte bis zum 12.07.2014. Es wurden Studierende der Hochschule Trier sowie der Universität Trier über den jeweiligen Emailverteiler kontaktiert. Insgesamt nahmen 1056 Personen an der Befragung teil, davon waren 37,3% weiblich; das Durchschnittsalter der Teilnehmer lag bei 27 Jahren.

Ergebnisse

Unsere Analyse fokussiert die Einflüsse auf die Qualität der Daten aus zwei unterschiedlichen Perspektiven. In einem ersten Schritt betrachten wir den Einfluss der verschiedenen Befragungsszenarien auf die Abbruchs- beziehungsweise Beendigungsquoten. In einem zweiten Schritt fokussieren wir die Einflüsse auf das Antwortverhalten.

Abbruchs- und Beendigungsquoten

Tab. 1 zeigt die Abbruch- und Beendigungsquote nach den verschiedenen Versionen. Dabei zeigt sich kein signifikanter Unterschied zwischen der Abbruchquote des gesamten Fragebogens und den verschiedenen Bedingungen. Die Abbruchquote ist über alle Versionen sehr niedrig und liegt zwischen 15,9% und 17,6%.

Tab. 1: Dropouts nach Version

	Version 1 (NFA)		Version 2 (FA)		Version 3 (FA/PNA)	
	N	%	N	%	N	%
Abgebrochen	56	15,9%	55	15,6%	62	17,6%
Beendet	296	84,1%	297	84,4%	290	82,4%
Gesamt	352		352		352	

Abb. 2 zeigt die kumulierten Abbruchhäufigkeiten der Befragung nach den verschiedenen Versionen. Hier konnten hochsignifikante Zusammenhänge bzgl. des Abbruchverlaufs festgestellt werden (Cramers $V=0,762^{***}$).

In der FA-Option sind die Abbrüche insbesondere am Anfang sehr hoch und zwar an einer Stelle im Fragebogen, an der den Befragten hintereinander zwei aufwändige Item-Batterien zu Uli Hoeneß präsentiert wurden. An dieser Stelle messen wir 87% aller Abbrüche in der FA-Option. Hier scheint sich das von Stieger, Reips und Voracec (2007) beschriebene Phänomen zu zeigen: Gering motivierte Befragte werden durch diese weniger spannenden Item-Batterien und durch den Zwang diese komplett zu beantworten zum Abbruch gebracht. In den anderen Versionen zeigt sich dieser Effekt nicht. Hier findet der Abbruch in beiden Versionen erst später statt und zwar zu dem Zeitpunkt, an dem die sensiblen Fragen gestellt werden (ohne FA 89%, FA/PNA 96%). Auf die Abbruchquote in der FA-Bedingung haben diese sensiblen Fragen jedoch nur noch einen geringen Effekt. Hier könnte durch die von uns in Aussicht gestellten Incentives ein »Behavioral Sunk Costs« Effekt (Cunha Jr, Caldieraro 2009; Zeelenberg, Van Dijk 1997) aufgetreten sein, denn die Befragten in der FA-Option haben zu diesem Zeitpunkt schon relativ viel Zeit investiert, da sie jede Frage beantworten mussten und lassen sich nun auch nicht durch die sensiblen Fragen zu einem Abbruch und damit zu dem Verlust der Chance ein Incentive zu gewinnen bringen.

Abb. 2: Abbruchverläufe während der Befragung nach Version



Über eine Einfaktorielle Varianzanalyse, die die Mittelwerte der kumulierten Abbruchhäufigkeiten⁶ nach den drei Versionen betrachtet, lassen sich die Unterschiede in den drei Versionen präzisieren. Während sich zwischen der Version ohne FA ($\bar{x}=20,91$) und der FA/PNA-Version ($\bar{x}=22,01$) keine relevanten Unterschiede feststellen lassen, hat die FA-Bedingung den höchsten Mittelwert ($\bar{x}=37,67$), wonach die Befragten dort am frühesten abbrechen.

Dementsprechend zeigen sich auch im Post-hoc-Test mittels Scheffé-Prozedur (Tab. 2) hochsignifikante Unterschiede zwischen den Versionen NFA und FA/PNA zur FA-Version. Vergleicht man die Abbruchverläufe zwischen der Forced Answering-Option mit Ausweichkategorie und derjenigen ohne Forced-Answering lassen sich keine signifikanten Unterschiede zwischen diesen beiden Versionen feststellen.

Tab. 2: Scheffé-Test

(I) Version	(J) Version	Mittlere Differenz (I-J)	Standardfehler	Signifikanz
NFA	FA	-16,764	2,31	0,000
	FA/PNA	-1,101	2,31	0,893
FA	NFA	16,764	2,31	0,000
	FA/PNA	15,663	2,31	0,000
FA/PNA	NFA	1,101	2,31	0,893
	FA	-15,663	2,31	0,000

Antwortverhalten

Tab. 3 bildet die Verteilung der Antworten auf die Frage nach eigenen Straftaten differenziert nach den Versionen ab. Wir haben diese als sehr sensibel einzuordnende Frage exemplarisch ausgewählt, um das jeweilige Antwortverhalten bei sehr persönlichen und heiklen Themen beschreiben zu können.

Es zeigt sich zum einen, dass die Befragten in der FA-Option signifikant seltener angeben, eine Straftat begangen zu haben. In beiden FA-Versionen sind es etwa 20 Prozent der Befragten, in der Version ohne Antwortzwang sind es hingegen 26 Prozent. Dies lässt – wie bei Stieger 2007 – einen Reaktanzeffekt vermuten. Dann würde der Antwortzwang dazu führen, dass die Tendenz sozial wünschenswert und dadurch nicht ehrlich zu antworten steigt. Weiterhin stellen wir fest, dass bei der FA-Version ohne neutrale Ausweichkategorie 80 Prozent mit »Nein« antworteten, während es bei der FA/PNA-Version 73 Prozent sind, die »Nein« und 6 Prozent, die die Kategorie »keine Angabe« wählten. Diese beiden Auffälligkeiten könnten zumindest ein Indiz für unsere dritte Hypothese sein: FA führt zu einer Abnahme valider Antworten. Um diese Vermutung jedoch adäquat zu begründen bedarf es weiteren Untersuchungen auf Basis von spezifischen Untersuchungssitem, die es ermöglichen, das Antwortverhalten gezielt zu analysieren.

⁶ Ein höherer Mittelwert bedeutet demnach, dass diese Fragebogenversion im Durchschnitt früher abgebrochen wurde, da die kumulierte Abbruchhäufigkeit schon an früherer Stelle relativ hoch war.

Tab. 3: Angabe zu eigenen Straftaten nach Version

Eigene Straftaten	Version 1 (NFA)	Version 2 (FA)	Version 3 (FA/PNA)
Ja	26,4%	19,5%	20,7%
Nein	72,3%	80,5%	73,1%
Keine Angabe	-	-	6,2%
nicht beantwortet	1,4%	-	-
Gesamt (beendet)	296	297	290

Sig.=.000, Cramer's V=.152

Zusammenfassung der Ergebnisse

Hinsichtlich der gesamten Dropout-Raten könnten wir in unserem Experiment keine Unterschiede zwischen den drei verschiedenen Bedingungen NFA, FA und FA/PNA feststellen. Die Abbruchverläufe und damit die Zeitpunkte des jeweiligen Abbruchs innerhalb der Befragung unterscheiden sich hingegen. Es wird deutlich, dass der Zwang auf Fragen zu antworten dazu führt, dass die Teilnehmer signifikant früher abbrechen. Dieser Effekt kann jedoch über eine neutrale Antwortkategorie deutlich abgeschwächt werden. Die fehlende Freiheit auf eine Frage nicht antworten zu müssen, sei es durch das Überspringen einer Frage oder durch eine »keine Angabe«-Funktion, scheint dazu zu führen, dass in der Forced Answering-Version bei heiklen Fragen häufiger die sozial erwünschte Antwortkategorie gewählt wird, beispielsweise um eine potenziell diskreditierende oder zu persönliche Information nicht preisgeben zu müssen. Entsprechend unserer Forschungshypothesen lässt sich Folgendes zusammenfassen:

- Zu Hypothese 1: FA hat in unserer Studie zwar nicht insgesamt zu einem erhöhten Unit-Nonresponse geführt, allerdings unterschieden sich die Zeitpunkte der Abbrüche signifikant.
- Zu Hypothese 2: Die Möglichkeit einer »keine Angabe«-Option hat diesen Effekt signifikant reduziert.
- Zu Hypothese 3: Die Antworten bei sensiblen Fragen unterscheiden sich zwischen den verschiedenen Bedingungen signifikant. Unter Antwortzwang scheinen die Befragten eher zu sozial wünschenswerten Antworten zu tendieren. Allerdings lässt sich anhand unserer Daten nicht eindeutig sagen, ob die Angaben in der NFA-Option tatsächlich valider sind.

Kritische Reflexion der Ergebnisse

Insgesamt fällt auf, dass die Abbruchquoten in allen drei Versionen der Befragungen vergleichsweise gering sind. Gründe hierfür gibt es mehrere: Zum einen handelt es sich bei Studierenden um eine Stichprobe mit ganz spezifischen Eigenschaften, beispielsweise einer hohen Affinität zu Umfragen (Jacob et al. 2013; Schnell et al. 2011). Weiterhin findet man bei einer Onlineumfrage eine höhere Bereitschaft Auskünfte, auch über sensible Themen zu geben (Décieux 2012; Dé-

cieux, Hoffmann 2014; Taddicken 2008). Ebenfalls ist zu vermuten, dass über den langen und einfachen Einstieg mit interessanten, kognitiv wenig anspruchsvollen und wenig sensiblen Fragen zum Fall Uli Hoeneß das Interesse an der Befragung und den noch folgenden Fragen geweckt wurde (Jacob et al. 2013; Meier et al. 2005). Einen besonders positiven Effekt auf die Abbruchquote werden jedoch die verhältnismäßig hohen Incentives gehabt haben (Göritz 2007). Bei der Bewertung und Interpretation des Effektes der Abbruchquoten sollte also wie auch schon bei den Studien von Albaum et. al (Albaum et al. 2014; Albaum et al. 2010; Albaum et al. 2011; Roster et al. 2014) bedacht werden, dass die Befragungssituation hier eine sehr spezifische ist, weil die Motivation die Umfrage zu beenden im Rahmen dieser Befragung deutlich höher ist, als unter realen Bedingungen.

Fazit

Auch wenn in diesem Experiment (unter Umständen aus oben genannten Gründen) kein Effekt auf die Höhe der Abbruchquoten insgesamt nachgewiesen wurde, so haben wir gezeigt, dass die Implementierung einer Forced Answering-Option in einer Onlinebefragung in mehrfacher Hinsicht einen Einfluss auf die Ergebnisse haben kann. Diese führt zum einen dazu, dass die Respondenten früher abbrechen beziehungsweise die neutrale Antwortkategorie häufiger wählen. Einen Effekt auf die Validität des Antwortverhaltens konnte nicht eindeutig nachgewiesen werden; es zeigt sich jedoch, dass unter Antwortzwang, signifikant häufiger die sozial wünschenswerte Antwortkategorie gewählt wurde.

Sowohl was die Analyse der Abbruchquoten als auch was die Validität der Antworten angeht, bedarf es weiterer Experimente, bei denen einerseits die Motivation eine Befragung zu beenden realistischer ist (weniger hohe oder keine Incentives) und die andererseits über die Implementierung von spezifischen Items und Fragelementen, fundierte Aussagen über die Zufälligkeit und die Ehrlichkeit der Antworten erlauben.

Insgesamt kommen wir bezüglich der Frage der Implementierung der Forced Answering-Option bei Onlinebefragungen zu dem Schluss, dass diese tendenziell eher einen negativen Einfluss auf die Qualität der Antworten und die Ergebnisse haben. Aus diesem Grund empfehlen wir diese nur in spezifischen Anwendungsfällen anzuwenden.

Literatur

- Albaum, G., Roster, C., Smith, S. 2014: Topic sensitivity and research design: effects on internet survey respondents' motives. *Asia Pacific Journal of Marketing and Logistics* 26, 147–161.
- Albaum, G., Wiley, J., Roster, C., Smith, S. M. 2011: Visiting item non-responses in internet survey data collection. *International Journal of Market Research* 53, 687–703.
- Albaum, G. et al. 2010: Designing web surveys in marketing research: does use of forced answering affect completion rates? *The Journal of Marketing Theory and Practice* 18, 285–294.
- Batinic, B. 2003: Internetbasierte Befragungsverfahren. *Österreichische Zeitschrift für Soziologie* 28, 6–18.

- Cunha Jr, M., Caldieraro, F. 2009: Sunk-Cost Effects on Purely Behavioral Investments. *Cognitive Science* 33, 105–113.
- Décieux, J. P. P. 2012: Modeeffekte bei Onlineumfragen: Ein multivariater Methodenvergleich unter Zuhilfenahme eines Propensity Score Matchings. München: Akademische Verlagsgemeinschaft.
- Décieux, J. P. P., Hoffmann, M. 2014: Antwortdifferenzen im Junk & Crime Survey: Ein Methodenvergleich mit goffmanscher Interpretation. S. In M. Löw (Hg.), Vielfalt und Zusammenhalt. Verhandlungen des 36. Kongresses der Deutschen Gesellschaft für Soziologie in Bochum und Dortmund 2012. Frankfurt am Main: Campus.
- Derouvray, C., Couper, M. P. 2002: Designing a strategy for reducing «no opinion» responses in web-based surveys. *Social Science Computer Review* 20, 3–9.
- Dillman, D. A., Smyth, J. D., Christian, L. M. 2014: Internet, phone, mail, and mixed-mode surveys: the tailored design method: John Wiley & Sons.
- Fan, W., Yan, Z. 2010: Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior* 26, 132–139.
- Galesic, M., Bosnjak, M. 2009: Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly* 73, 349–360.
- Ganassali, S. 2008: The influence of the design of web survey questionnaires on the quality of responses. *Survey Research Methods*.
- Göritz, A. S. 2007: Using online panels in psychological research. *The Oxford handbook of Internet psychology*, 473–485.
- Jacob, R., Heinz, A., Décieux, J. P. 2013: Umfrage: Einführung in die Methoden der Umfrageforschung. München: Oldenbourg Verlag.
- Kaczmirek, L. 2005: Web Surveys: A Brief Guide on Usability and Implementation Issues. Center for Survey Research and Methodology (ZUMA) 22.
- Marcus, B. et al. 2007: Compensating for Low Topic Interest and Long Surveys: A Field Experiment on Non-response in Web Surveys. *Social Science Computer Review* 25, 372–383.
- Meier, G., Schneid, M., Stegemann, Y., Stiegler, A. 2005: Steigerung der Ausschöpfungsquote von Telefonumfragen durch geschickte Einleitungstexte. *Zuma Nachrichten* 29: 37–55.
- O'Neil, K. M., Penrod, S. D., Bornstein, B. H. 2003: Web-based research: Methodological variables' effects on dropout and sample characteristics. *Behavior Research Methods, Instruments, & Computers* 35, 217–226.
- Peytchev, A., Crawford, S. 2005: A typology of real-time validations in web-based surveys. *Social science computer review* 23, 235–249.
- Preisendörfer, P. 2008: Heikle Fragen in mündlichen Interviews: Ergebnisse einer Methodenstudie im studentischen Milieu. Zürich. Working Paper. 6.
- Questback 2013: Enterprise Feedback Suite EFS Survey 9.1. Köln.
- Roster, C. A., Albaum, G., Smith, S. M. 2014: Topic sensitivity and Internet survey design: A cross-cultural/national study. *Journal of Marketing Theory and Practice* 22, 91–102.
- Sakshaug, J. W., Yan, T., Tourangeau, R. 2010: Nonresponse error, measurement error, and mode of data collection: tradeoffs in a multi-mode survey of sensitive and non-sensitive items. *Public Opinion Quarterly* 74, 907–933.
- Schnell, R., Hill, P. B., Esser, E. 2011: Methoden der empirischen Sozialforschung. München: Oldenbourg Verlag.
- Stieger, S., Reips, U. D., Voracek, M. 2007: Forced-response in online surveys: Bias from reactance and an increase in sex-specific dropout. *Journal of the American Society for Information Science and Technology* 58, 1653–1660.
- Taddicken, M. 2008: Methodeneffekte bei Web-Befragungen. Köln: Herbert von Halem Verlag.
- Vicente, P., Reis, E. 2010: Using questionnaire design to fight nonresponse bias in web surveys. *Social Science Computer Review* 28, 251–267.
- Zeelenberg, M., Van Dijk, E. 1997: A reverse sunk cost effect in risky decision making: Sometimes we have too much invested to gamble. *Journal of Economic Psychology* 18, 677–691.